PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C12N 15/11, 15/12, C07K 14/705, A61K 38/17, C12Q 1/68, G01N 33/68, C07K 16/30

A2

(11) International Publication Number: **WO 00/58460**

(43) International Publication Date: 5 October 2000 (05.10.00)

(21) International Application Number: PCT/EP00/02478

(22) International Filing Date: 20 March 2000 (20.03.00)

(30) Priority Data:

9907113.6	26 March 1999 (26.03.99)	GB
9922858.7	25 September 1999 (25.09.99)	GB

(71) Applicant *(for all designated States except US)*: SMITHKLINE BEECHAM BIOLOGICALS S.A. [BE/BE]; Rue de l'Institut 89, B–1330 Rixensart (BE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: BRUCK, Claudine, Elvire, Marie [BE/BE]; SmithKline Beecham Biologicals S.A., Rue de l'Institut 89, B–1330 Rixensart (BE). CASSART, Jean–Pol [BE/BE]; SmithKline Beecham Biologicals S.A., Rue de l'Institut 89, B–1330 Rixensart (BE). COCHE, Thierry [BE/BE]; SmithKline Beecham Biologicals S.A., Rue de l'Institut 89, B–1330 Rixensart (BE). VINALS Y DE BASSOLS, Carlota [BE/BE]; SmithKline Beecham Biologicals S.A., Rue de l'Institut 89, B–1330 Rixensart (BE).

(74) Agent: PRIVETT, Kathryn, Louise; SmithKline Beecham, Two New Horizons Court, Brentford, Middlesex TW8 9EP (GB).

(81) Designated States: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: NOVEL COMPOUNDS

(57) Abstract

CASB619 polypeptides and polynucleotides and methods for producing such polypeptides by recombinant techniques are disclosed. Also disclosed are methods for utilizing CASB619 polypeptides and polynucleotides in diagnostics, and vaccines for prophylactic and therapeutic treatment of cancers, particularly ovarian and colon cancers, autoimmune diseases, and related conditions.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

N vel Compounds

The present invention relates to polynucleotides, herein referred to as CASB619 polynucleotides, polypeptides encoded thereby (referred to herein as CASB619 polypeptides), recombinant materials and methods for their production. In another aspect, the invention relates to methods for using such polypeptides and polynucleotides, including the treatment of cancer, in particular colon cancer and autoimmune diseases and other related conditions. In a further aspect, the invention relates to methods for identifying agonists and antagonists/inhibitors using the materials provided by the invention, and treating conditions associated with CASB619 polypeptide imbalance with the identified compounds. In a still further aspect, the invention relates to diagnostic assays for detecting diseases associated with inappropriate CASB619 polypeptide activity or levels.

Polypeptides and polynucleotides of the present invention are believed to be important immunogens for specific prophylactic or therapeutic immunization against tumours, because they are specifically expressed or highly over-expressed in tumours compared to normal cells and can thus be targeted by antigen-specific immune mechanisms leading to the destruction of the tumour cell. They can also be used to diagnose the occurrence of tumour cells. Furthermore, their inappropriate expression in certain circumstances can cause an induction of autoimmune, inappropriate immune responses, which could be corrected through appropriate vaccination using the same polypeptides or polynucleotides. In this respect the most important biological activities to our purpose are the antigenic and immunogenic activities of the polypeptide of the present invention. A polypeptide of the present invention may also exhibit at least one other biological activity of a CASB619 polypeptide, which could qualify it as a target for therapeutic or prophylactic intervention different from that linked to the immune response.

In a first aspect, the present invention relates to CASB619 polypeptides. Such peptides include isolated polypeptides comprising an amino acid sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to that of SEQ ID NO:2 over the entire length of SEQ ID NO:2. Such polypeptides include those comprising the amino acid of SEQ ID NO:2.

Further peptides of the present invention include isolated polypeptides in which the amino acid sequence has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to the amino acid sequence of SEQ ID NO:2 over the entire length of SEQ ID NO:2. Such polypeptides include the polypeptide of SEQ ID NO:2.

Further peptides of the present invention include isolated polypeptides encoded by a polynucleotide comprising the sequence contained in SEQ ID NO:1.

The invention also provides an immunogenic fragment of a CASB619 polypeptide, that is a contiguous portion of the CASB619 polypeptide which has the same or similar immunogenic properties to the polypeptide comprising the amino acid seqeunce of SEQ ID NO:2. That is to say, the fragment (if necessary when coupled to a carrier) is capable of raising an immune response which recognises the CASB619 polypeptide. Such an immunogenic fragment may include, for example, the CASB619 polypeptide lacking an N-terminal leader sequence, a transmembrane domain or a C-terminal anchor domain. In a preferred aspect the immunogenic fragment of CASB619 according to the invention comprises substantially all of the extracellular domain of a polypeptide which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to that of SEQ ID NO:2 over the entire length of SEQ ID NO:2.

Peptide fragments incorporating an epitope of CASB619 typically will comprise at least 7, preferably 9 or 10 contiguous amino acids from SEQ ID NO:2. Preferred epitopes are shown in SEQ ID NO:5 to SEQ ID NO:68.

Peptides that incorporate these epitopes form a preferred aspect of the present invention. Mimotopes which have the same characteristics as these epitopes, and immunogens comprising such mimotopes which generate an immune response which cross-react with an epitope in the context of the CASB619 molecule, also form part of the present invention.

The present invention, therefore, includes isolated peptides encompassing these epitopes themselves, and any mimotope thereof. The meaning of mimotope is defined as an entity which is sufficiently similar to the native CASB619 epitope so as to be capable of being recognised by antibodies which recognise the native molecule; (Gheysen, H.M., et al., 1986, Synthetic peptides as antigens. Wiley, Chichester, Ciba foundation symposium 119, p130-149; Gheysen, H.M., 1986, Molecular Immunology, 23,7, 709-715); or are capable of raising antibodies, when coupled to a suitable carrier, which antibodies cross-react with the native molecule.

Peptide mimotopes of the above-identified epitopes may be designed for a particular purpose by addition, deletion or substitution of elected amino acids. Thus, the peptides of the present invention may be modified for the purposes of ease of conjugation to a protein carrier. For example, it may be desirable for some chemical conjugation methods to include a terminal cysteine to the epitope. In addition it may be desirable for peptides conjugated to a protein carrier to include a hydrophobic terminus distal from the conjugated terminus of the peptide, such that the free unconjugated end of the peptide remains associated with the surface of the carrier protein. This reduces the conformational degrees of freedom of the peptide, and thus increases the probability that the peptide is presented in a conformation which most closely resembles that of the peptide as found in the context of the whole molecule. For example, the peptides may be altered to have an N-terminal cysteine and a C-terminal hydrophobic amidated tail. Alternatively, the addition or substitution of a D-stereoisomer form of one or more of the amino acids may be performed to create a beneficial derivative, for example to enhance stability of the peptide. Those skilled in the art will realise that such modified peptides, or mimotopes, could be a wholly or partly non-peptide mimotope wherein the constituent residues are not necessarily confined to the 20 naturally occurring amino acids. In addition, these may be cyclised by techniques known in the art to constrain the peptide into a conformation that closely resembles its shape when the peptide sequence is in the context of the whole molecule. A preferred method of cyclising a peptide comprises the addition of a pair of cysteine residues to allow the formation of a disulphide bridge.

Further, those skilled in the art will realise that mimotopes or immunogens of the present invention may be larger than the above-identified epitopes, and as such may comprise the sequences disclosed herein. Accordingly, the mimotopes of the present invention may consist of addition of N and/or C terminal extensions of a number of other natural residues at one or both ends. The peptide mimotopes may also be retro sequences of the natural sequences, in that the sequence orientation is reversed; or alternatively the sequences may be entirely or at least in part comprised of D-stereo isomer amino acids (inverso sequences). Also, the peptide sequences may be retro-inverso in character, in that the sequence orientation is reversed and the amino acids are of the D-stereoisomer form. Such retro or retro-inverso peptides have the advantage of being non-self, and as such may overcome problems of self-tolerance in the immune system.

Alternatively, peptide mimotopes may be identified using antibodies which are capable themselves of binding to the epitopes of the present invention using techniques such as phage display technology (EP 0 552 267 B1). This technique, generates a large number of peptide sequences which mimic the structure of the native peptides and are, therefore, capable of binding to anti-native peptide antibodies, but may not necessarily themselves share significant sequence homology to the native peptide. This approach may have significant advantages by allowing the possibility of identifying a peptide with enhanced immunogenic properties, or may overcome any potential self-antigen tolerance problems which may be associated with the use of the native peptide sequence. Additionally this technique allows the identification of a recognition pattern for each native-peptide in terms of its shared chemical properties amongst recognised mimotope sequences.

The covalent coupling of the peptide to the immunogenic carrier can be carried out in a manner well known in the art. Thus, for example, for direct covalent coupling it is possible to utilise a carbodiimide, glutaraldehyde or (N-[γ-maleimidobutyryloxy] succinimide ester, utilising common commercially available heterobifunctional linkers such as CDAP and SPDP (using manufacturers instructions). After the coupling reaction, the immunogen can easily be isolated and purified by means of a dialysis method, a gel filtration method, a fractionation method etc.

The types of carriers used in the immunogens of the present invention will be readily known to the man skilled in the art. The function of the carrier is to provide cytokine help in order to help induce an immune response against the peptide. A non-exhaustive list of carriers which may be used in the present invention include: Keyhole limpet Haemocyanin (KLH), serum albumins such as bovine serum albumin (BSA), inactivated bacterial toxins such as tetanus or diptheria toxins (TT and DT), or recombinant fragments thereof (for example, Domain 1 of Fragment C of TT, or the translocation domain of DT), or the purified protein derivative of tuberculin (PPD). Alternatively the mimotopes or epitopes may be directly conjugated to liposome carriers, which may additionally comprise immunogens capable of providing T-cell help. Preferably the ratio of mimotopes to carrier is in the order of 1:1 to 20:1, and preferably each carrier should carry between 3-15 peptides.

In an embodiment of the invention a preferred carrier is Protein D from *Haemophilus influenzae* (EP 0 594 610 B1). Protein D is an IgD–binding protein from *Haemophilus influenzae* and has been patented by Forsgren (WO 91/18926, granted EP 0 594 610 B1). In some circumstances, for example in recombinant immunogen expression systems it may be desirable to use fragments of protein D, for example Protein D 1/3rd (comprising the N-terminal 100-110 amino acids of protein D (GB 9717953.5)).

Another preferred method of presenting the peptides of the present invention is in the context of a recombinant fusion molecule. For example, EP 0 421 635 B describes the use of chimaeric hepadnavirus core antigen particles to present foreign peptide sequences in a virus-like particle. As such, immunogens of the present invention may comprise peptides presented in chimaeric particles consisting of hepatitis B core antigen. Additionally, the recombinant fusion proteins may comprise the mimotopes of the present invention and a carrier protein, such as NS1 of the influenza virus. For any recombinantly expressed protein which forms part of the present invention, the nucleic acid which encodes said immunogen also forms an aspect of the present invention.

Peptides used in the present invention can be readily synthesised by solid phase procedures well known in the art. Suitable syntheses may be performed by utilising "T-boc" or "F-moc" procedures. Cyclic peptides can be synthesised by the solid phase

procedure employing the well-known "F-moc" procedure and polyamide resin in the fully automated apparatus. Alternatively, those skilled in the art will know the necessary laboratory procedures to perform the process manually. Techniques and procedures for solid phase synthesis are described in 'Solid Phase Peptide Synthesis: A Practical Approach' by E. Atherton and R.C. Sheppard, published by IRL at Oxford University Press (1989). Alternatively, the peptides may be produced by recombinant methods, including expressing nucleic acid molecules encoding the mimotopes in a bacterial or mammalian cell line, followed by purification of the expressed mimotope. Techniques for recombinant expression of peptides and proteins are known in the art, and are described in Maniatis, T., Fritsch, E.F. and Sambrook et al., *Molecular cloning, a laboratory manual*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989).

The polypeptides or immunogenic fragment of the invention may be in the form of the "mature" protein or may be a part of a larger protein such as a precursor or a fusion protein. It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification such as multiple histidine residues, or an additional sequence for stability during recombinant production. Furthermore, addition of exogenous polypeptide or lipid tail or polynucleotide sequences to increase the immunogenic potential of the final molecule is also considered.

In one aspect, the invention relates to genetically engineered soluble fusion proteins comprising a polypeptide of the present invention, or a fragment thereof, and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclasses (IgG, IgM, IgA, IgE). Preferred as an immunoglobulin is the constant part of the heavy chain of human IgG, particularly IgG1, where fusion takes place at the hinge region. In a particular embodiment, the Fc part can be removed simply by incorporation of a cleavage sequence which can be cleaved with blood clotting factor Xa. Furthermore, this invention relates to processes for the preparation of these fusion proteins by genetic engineering, and to the use thereof for drug screening, diagnosis and therapy. A further aspect of the invention also relates to polynucleotides encoding such fusion proteins.

Examples of fusion protein technology can be found in International Patent Application Nos. WO94/29458 and WO94/22914.

The proteins may be chemically conjugated, or expressed as recombinant fusion proteins allowing increased levels to be produced in an expression system as compared to non-fused protein. The fusion partner may assist in providing T helper epitopes (immunological fusion partner), preferably T helper epitopes recognised by humans, or assist in expressing the protein (expression enhancer) at higher yields than the native recombinant protein. Preferably the fusion partner will be both an immunological fusion partner and expression enhancing partner.

Fusion partners include protein D from *Haemophilus influenza* B and the non-structural protein from influenzae virus, NS1 (hemagglutinin). Another immunological fusion partner is the protein known as LYTA. Preferably the C terminal portion of the molecule is used. Lyta is derived from *Streptococcus pneumoniae* which synthesize an N-acetyl-L-alanine amidase, amidase LYTA, (coded by the lytA gene {Gene, 43 (1986) page 265-272} an autolysin that specifically degrades certain bonds in the peptidoglycan backbone. The C-terminal domain of the LYTA protein is responsible for the affinity to the choline or to some choline analogues such as DEAE. This property has been exploited for the development of E.coli C-LYTA expressing plasmids useful for expression of fusion proteins. Purification of hybrid proteins containing the C-LYTA fragment at its amino terminus has been described {Biotechnology: 10, (1992) page 795-798}. It is possible to use the repeat portion of the Lyta molecule found in the C terminal end starting at residue 178, for example residues 188 - 305.

The present invention also includes variants of the aforementioned polypeptides, that is polypeptides that vary from the referents by conservative amino acid substitutions, whereby a residue is substituted by another with like characteristics. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; and among the basic residues Lys and Arg; or aromatic residues Phe and Tyr. Particularly preferred are variants in which several, 5-10, 1-5, 1-3, 1-2 or 1 amino acids are substituted, deleted, or added in any combination.

Polypeptides of the present invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

In a further aspect, the present invention relates to CASB619 polynucleotides. Such polynucleotides include isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, to the amino acid sequence of SEQ ID NO:2, over the entire length of SEQ ID NO:2. In this regard, polypeptides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identity are more highly preferred, and those with at least 99% identity are most highly preferred. Such polynucleotides include a polynucleotide comprising the nucleotide sequence contained in SEQ ID NO:1 encoding the polypeptide of SEQ ID NO:2.

Further polynucleotides of the present invention include isolated polynucleotides comprising a nucleotide sequence that has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, to a nucleotide sequence encoding a polypeptide of SEQ ID NO:2, over the entire coding region. In this regard, polynucleotides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identity are more highly preferred, and those with at least 99% identity are most highly preferred.

Further polynucleotides of the present invention include isolated polynucleotides comprising a nucleotide sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, to SEQ ID NO:1 over the entire length of SEQ ID NO:1. In this regard, polynucleotides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identiy are more highly preferred, and those with at least 99% identity are most highly preferred. Such polynucleotides include a polynucleotide comprising the polynucleotide of SEQ ID NO:1 as well as the polynucleotide of SEQ ID NO:1. Said polynucleotide can be inserted in a suitable plasmid or recombinant microrganism vector and used for

immunization (see for example Wolff et. al., Science 247:1465-1468 (1990); Corr et. al., J. Exp. Med. 184:1555-1560 (1996); Doe et. al., Proc. Natl. Acad. Sci. 93:8578-8583 (1996)). The invention also provides polynucleotides which are complementary to all the above described polynucleotides.

The invention also provides a fragment of a CASB619 polynucleotide which when administered to a subject has the same immunogenic properties as the polynucleotide of SEQ ID NO:1.

The invention also provides a polynucleotide encoding an immunological fragment of a CASB619 polypeptide as hereinbefore defined.

The nucleotide sequence of SEQ ID NO:1 shows homology with Homo sapiens chromosome 1 clone RP4-641D22 map p13.1-13.3 (accession AL157901). The nucleotide sequence of SEQ ID NO:1 is a cDNA sequence and comprises a polypeptide encoding sequence (nucleotide 1 to 3043) encoding a polypeptide of 1013 amino acids, the polypeptide of SEQ ID NO:2. The nucleotide sequence encoding the polypeptide of SEQ ID NO:2 may be identical to the polypeptide encoding sequence contained in SEQ ID NO:1 or it may be a sequence other than the one contained in SEQ ID NO:1, which, as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide of SEQ ID NO:2. The polypeptide of the SEQ ID NO:2 is notrelated to any other known protein.

Preferred polypeptides and polynucleotides of the present invention are expected to have, *inter alia*, similar biological functions/properties to their homologous polypeptides and polynucleotides. Furthermore, preferred polypeptides, immunological fragments and polynucleotides of the present invention have at least one activity of either SEQ ID NO:1 or SEQ ID NO:2, as appropriate.

The present invention also relates to partial or other incomplete polynucleotide and polypeptide sequences which were first identified prior to the determination of the corresponding full length sequences of SEQ ID NO:1 and SEQ ID NO:2.

Accordingly, in a further aspect, the present invention provides for an isolated polynucleotide which:

(a) comprises a nucleotide sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, even more preferably at least 97-99% identity to SEQ ID NO:3 over the entire length of SEQ ID NO:3;

(b) has a nucleotide sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, even more preferably at least 97-99% identity, to SEQ ID NO:1 over the entire length of SEQ ID NO:3;

(c) the polynucleotide of SEQ ID NO:3; or

(d) a nucleotide sequence encoding a polypeptide which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, even more preferably at least 97-99% identity, to the amino acid sequence of SEQ ID NO:4, over the entire length of SEQ ID NO:4;

as well as the polynucleotide of SEQ ID NO:3.

The present invention further provides for a polypeptide which:

(a) comprises an amino acid sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to that of SEQ ID NO:2 over the entire length of SEQ ID NO:4;

(b) has an amino acid sequence which is at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to the amino acid sequence of SEQ ID NO:2 over the entire length of SEQ ID NO:4;

(c) comprises the amino acid of SEQ ID NO:4; and

(d) is the polypeptide of SEQ ID NO:4;

as well as polypeptides encoded by a polynucleotide comprising the sequence contained in SEQ ID NO:3.

The nucleotide sequence of SEQ ID NO:3 and the peptide sequence encoded thereby are derived from EST (Expressed Sequence Tag) sequences. It is recognised by those skilled

in the art that there will inevitably be some nucleotide sequence reading errors in EST sequences (see Adams, M.D. *et al*, Nature 377 (supp) 3, 1995). Accordingly, the nucleotide sequence of SEQ ID NO:3 and the peptide sequence encoded therefrom are therefore subject to the same inherent limitations in sequence accuracy. Furthermore, the peptide sequence encoded by SEQ ID NO:3 comprises a region of identity or close homology and/or close structural similarity (for example a conservative amino acid difference) with the closest homologous or structurally similar protein.

Polynucleotides of the present invention may be obtained, using standard cloning and screening techniques, from a cDNA library derived from mRNA in cells of human colon cancer, (for example Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring harbor Laboratory Press, Cold Spring harbor, N.Y. (1989)). Polynucleotides of the invention can also be obtained from natural sources such as genomic DNA libraries or can be synthesized using well known and commercially available techniques.

When polynucleotides of the present invention are used for the recombinant production of polypeptides of the present invention, the polynucleotide may include the coding sequence for the mature polypeptide, by itself; or the coding sequence for the mature polypeptide in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro- protein sequence, or other fusion peptide portions. For example, a marker sequence which facilitates purification of the fused polypeptide can be encoded. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, Proc Natl Acad Sci USA (1989) 86:821-824, or is an HA tag. The polynucleotide may also contain non-coding 5' and 3' sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and sequences that stabilize mRNA.

Further embodiments of the present invention include polynucleotides encoding polypeptide variants which comprise the amino acid sequence of SEQ ID NO:2 and in which several, for instance from 5 to 10, 1 to 5, 1 to 3, 1 to 2 or 1, amino acid residues are substituted, deleted or added, in any combination.

Polynucleotides which are identical or sufficiently identical to a nucleotide sequence contained in SEQ ID NO:1, may be used as hybridization probes for cDNA and genomic DNA or as primers for a nucleic acid amplification (PCR) reaction, to isolate full-length cDNAs and genomic clones encoding polypeptides of the present invention and to isolate cDNA and genomic clones of other genes (including genes encoding paralogs from human sources and orthologs and paralogs from species other than human) that have a high sequence similarity to SEQ ID NO:1. Typically these nucleotide sequences are 70% identical, preferably 80% identical, more preferably 90% identical, most preferably 95% identical to that of the referent. The probes or primers will generally comprise at least 15 nucleotides, preferably, at least 30 nucleotides and may have at least 50 nucleotides. Particularly preferred probes will have between 30 and 50 nucleotides. Particularly preferred primers will have between 20 and 25 nucleotides. In particular, polypeptides or polynucleotides derived from sequences from homologous animal origin could be used as immunogens to obtain a cross-reactive immune response to the human gene.

A polynucleotide encoding a polypeptide of the present invention, including homologs from species other than human, may be obtained by a process which comprises the steps of screening an appropriate library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO: 1 or a fragment thereof; and isolating full-length cDNA and genomic clones containing said polynucleotide sequence. Such hybridization techniques are well known to the skilled artisan. Preferred stringent hybridization conditions include overnight incubation at 42°C in a solution comprising: 50% formamide, 5xSSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10 % dextran sulfate, and 20 microgram/ml denatured, sheared salmon sperm DNA; followed by washing the filters in 0.1x SSC at about 65°C. Thus the present invention also includes polynucleotides obtainable by screening an appropriate library under stingent hybridization conditions with a labeled probe having the sequence of SEQ ID NO:1 or a fragment thereof.

The skilled artisan will appreciate that, in many cases, an isolated cDNA sequence will be incomplete, in that the region coding for the polypeptide is short at the 5' end of the cDNA.

There are several methods available and well known to those skilled in the art to obtain full-length cDNAs, or extend short cDNAs, for example those based on the method of Rapid Amplification of cDNA ends (RACE) (see, for example, Frohman et al., PNAS USA 85, 8998-9002, 1988). Recent modifications of the technique, exemplified by the Marathon™ technology (Clontech Laboratories Inc.) for example, have significantly simplified the search for longer cDNAs. In the Marathon™ technology, cDNAs have been prepared from mRNA extracted from a chosen tissue and an 'adaptor' sequence ligated onto each end. Nucleic acid amplification (PCR) is then carried out to amplify the 'missing' 5' end of the cDNA using a combination of gene specific and adaptor specific oligonucleotide primers. The PCR reaction is then repeated using 'nested' primers, that is, primers designed to anneal within the amplified product (typically an adaptor specific primer that anneals further 3' in the adaptor sequence and a gene specific primer that anneals further 5' in the known gene sequence). The products of this reaction can then be analysed by DNA sequencing and a full-length cDNA constructed either by joining the product directly to the existing cDNA to give a complete sequence, or carrying out a separate full-length PCR using the new sequence information for the design of the 5' primer.

Recombinant polypeptides of the present invention may be prepared by processes well known in the art from genetically engineered host cells comprising expression systems. Accordingly, in a further aspect, the present invention relates to an expression system which comprises a polynucleotide of the present invention, to host cells which are genetically engineered with such expression sytems and to the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof for polynucleotides of the present invention. Introduction of polynucleotides into host cells can be effected by methods described in many standard laboratory manuals, such as Davis et al., Basic Methods in Molecular Biology (1986) and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold

Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989). Preferred such methods include, for instance, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection.

Preferably the proteins of the invention are coexpressed with thioredoxin in trans (TIT). Coexpression of thioredoxin in trans versus in cis is preferred to keep antigen free of thioredoxin without the need for protease. Thioredoxin coexpression eases the solubilisation of the proteins of the invention. Thioredoxin coexpression has also a significant impact on protein purification yield, on purified-protein solubility and quality.

Representative examples of appropriate hosts include bacterial cells, such as *Streptococci*, *Staphylococci*, *E. coli*, *Streptomyces* and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, HEK 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used, for instance, chromosomal, episomal and virus-derived systems, e.g., vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression systems may contain control regions that regulate as well as engender expression. Generally, any system or vector which is able to maintain, propagate or express a polynucleotide to produce a polypeptide in a host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, Molecular Cloning, A Laboratory Manual (supra). Appropriate secretion signals may be incorporated into the desired polypeptide to allow secretion of the translated protein into the lumen of the endoplasmic reticulum, the periplasmic space or the extracellular

environment. These signals may be endogenous to the polypeptide or they may be heterologous signals.

The expression system may also be a recombinant live microorganism, such as a virus or bacterium. The gene of interest can be inserted into the genome of a live recombinant virus or bacterium. Inoculation and *in vivo* infection with this live vector will lead to *in vivo* expression of the antigen and induction of immune responses. Viruses and bacteria used for this purpose are for instance: poxviruses (e.g; vaccinia, fowlpox, canarypox), alphaviruses (Sindbis virus, Semliki Forest Virus, Venezuelian Equine Encephalitis Virus), adenoviruses, adeno-associated virus, picornaviruses (poliovirus, rhinovirus), herpesviruses (varicella zoster virus, etc), Listeria, Salmonella , Shigella, BCG. These viruses and bacteria can be virulent, or attenuated in various ways in order to obtain live vaccines. Such live vaccines also form part of the invention.

Polypeptides of the present invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, ion metal affinity chromatography (IMAC) is employed for purification. Well known techniques for refolding proteins may be employed to regenerate active conformation when the polypeptide is denatured during intracellular synthesis, isolation and or purification.

Another important aspect of the invention relates to a method for inducing , re-inforcing or modulating an immunological response in a mammal which comprises inoculating the mammal with a fragment or the entire polypeptide or polynucleotide of the invention, adequate to produce antibody and/or T cell immune response for prophylaxis or for therapeutic treatment of cancer and autoimmune disease and related conditions. Yet another aspect of the invention relates to a method of inducing, re-inforcing or modulating immunological response in a mammal which comprises, delivering a polypeptide of the present invention *via* a vector or cell directing expression of the polynucleotide and coding for the polypeptide *in vivo* in order to induce such an

immunological response to produce immune responses for prophylaxis or treatment of said mammal from diseases.

A further aspect of the invention relates to an immunological/vaccine formulation (composition) which, when introduced into a mammalian host, induces, re-inforces or modulates an immunological response in that mammal to a polypeptide of the present invention wherein the composition comprises a polypeptide or polynucleotide of the invention or an immunological fragment thereof as herein before defined.The vaccine formulation may further comprise a suitable carrier. Since a polypeptide may be broken down in the stomach, it is preferably administered parenterally (for instance, subcutaneous, intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use.

A further aspect of the invention relates to the *in vitro* induction of immune responses to a fragment or the entire polypeptide or polynucleotide of the present invention or a molecule comprising the polypeptide or polynucleotide of the present invention, using cells from the immune system of a mammal, and reinfusing these activated immune cells of the mammal for the treatment of disease. Activation of the cells from the immune system is achieved by in vitro incubation with the entire polypeptide or polynucleotide of the present invention or a molecule comprising the polypeptide or polynucleotide of the present invention in the presence or absence of various immunomodulator molecules. A further aspect of the invention relates to the immunization of a mammal by administration of antigen presenting cells modified by *in vitro* loading with part or the entire polypeptide of the present invention or a molecule comprising the polypeptide of the present invention and administered *in vivo* in an immunogenic way. Alternatively, antigen presenting cells can be transfected *in vitro* with a vector containing a fragment or the entire polynucleotide of the present invention or a molecule comprising the

polynucleotide of the present invention, such as to express the corresponding polypeptide, and administered *in vivo* in an immunogenic way.

The vaccine formulation of the invention may also include adjuvant systems for enhancing the immunogenicity of the formulation. Preferably the adjuvant system raises preferentially a TH1 type of response.

An immune response may be broadly distinguished into two extreme catagories, being a humoral or cell mediated immune responses (traditionally characterised by antibody and cellular effector mechanisms of protection respectively). These categories of response have been termed TH1-type responses (cell-mediated response), and TH2-type immune responses (humoral response).

Extreme TH1-type immune responses may be characterised by the generation of antigen specific, haplotype restricted cytotoxic T lymphocytes, and natural killer cell responses. In mice TH1-type responses are often characterised by the generation of antibodies of the IgG2a subtype, whilst in the human these correspond to IgG1 type antibodies. TH2-type immune responses are characterised by the generation of a broad range of immunoglobulin isotypes including in mice IgG1, IgA, and IgM.

It can be considered that the driving force behind the development of these two types of immune responses are cytokines. High levels of TH1-type cytokines tend to favour the induction of cell mediated immune responses to the given antigen, whilst high levels of TH2-type cytokines tend to favour the induction of humoral immune responses to the antigen.

The distinction of TH1 and TH2-type immune responses is not absolute. In reality an individual will support an immune response which is described as being predominantly TH1 or predominantly TH2. However, it is often convenient to consider the families of cytokines in terms of that described in murine CD4 +ve T cell clones by Mosmann and Coffman (*Mosmann, T.R. and Coffman, R.L. (1989) TH1 and TH2 cells: different patterns of lymphokine secretion lead to different functional properties. Annual Review of Immunology, 7, p145-173*). Traditionally, TH1-type responses are associated with the

production of the INF-γ and IL-2 cytokines by T-lymphocytes. Other cytokines often directly associated with the induction of TH1-type immune responses are not produced by T-cells, such as IL-12. In contrast, TH2- type responses are associated with the secretion of IL-4, IL-5, IL-6 and IL-13.

It is known that certain vaccine adjuvants are particularly suited to the stimulation of either TH1 or TH2 - type cytokine responses. Traditionally the best indicators of the TH1:TH2 balance of the immune response after a vaccination or infection includes direct measurement of the production of TH1 or TH2 cytokines by T lymphocytes *in vitro* after restimulation with antigen, and/or the measurement of the IgG1:IgG2a ratio of antigen specific antibody responses.

Thus, a TH1-type adjuvant is one which preferentially stimulates isolated T-cell populations to produce high levels of TH1-type cytokines when re-stimulated with antigen *in vitro*, and promotes development of both CD8+ cytotoxic T lymphocytes and antigen specific immunoglobulin responses associated with TH1-type isotype.

Adjuvants which are capable of preferential stimulation of the TH1 cell response are described in International Patent Application No. WO 94/00153 and WO 95/17209.

3 De-O-acylated monophosphoryl lipid A (3D-MPL) is one such adjuvant. This is known from GB 2220211 (Ribi). Chemically it is a mixture of 3 De-O-acylated monophosphoryl lipid A with 4, 5 or 6 acylated chains and is manufactured by Ribi Immunochem, Montana. A preferred form of 3 De-O-acylated monophosphoryl lipid A is disclosed in European Patent 0 689 454 B1 (SmithKline Beecham Biologicals SA).

Preferably, the particles of 3D-MPL are small enough to be sterile filtered through a 0.22micron membrane (European Patent number 0 689 454).
3D-MPL will be present in the range of 10μg - 100μg preferably 25-50μg per dose wherein the antigen will typically be present in a range 2-50μg per dose.

Another preferred adjuvant comprises QS21, an Hplc purified non-toxic fraction derived from the bark of Quillaja Saponaria Molina. Optionally this may be admixed with 3 De-O-acylated monophosphoryl lipid A (3D-MPL), optionally together with an carrier.

The method of production of QS21 is disclosed in US patent No. 5,057,540.

Non-reactogenic adjuvant formulations containing QS21 have been described previously (WO 96/33739). Such formulations comprising QS21 and cholesterol have been shown to be successful TH1 stimulating adjuvants when formulated together with an antigen.

Further adjuvants which are preferential stimulators of TH1 cell response include immunomodulatory oligonucleotides, for example unmethylated CpG sequences as disclosed in WO 96/02555.

Combinations of different TH1 stimulating adjuvants, such as those mentioned hereinabove, are also contemplated as providing an adjuvant which is a preferential stimulator of TH1 cell response. For example, QS21 can be formulated together with 3D-MPL. The ratio of QS21 : 3D-MPL will typically be in the order of 1 : 10 to 10 : 1; preferably 1:5 to 5 : 1 and often substantially 1 : 1. The preferred range for optimal synergy is 2.5 : 1 to 1 : 1 3D-MPL: QS21.

Preferably a carrier is also present in the vaccine composition according to the invention. The carrier may be an oil in water emulsion, or an aluminium salt, such as aluminium phosphate or aluminium hydroxide.

A preferred oil-in-water emulsion comprises a metabolisible oil, such as squalene, alpha tocopherol and Tween 80. In a particularly preferred aspect the antigens in the vaccine composition according to the invention are combined with QS21 and 3D-MPL in such an emulsion. Additionally the oil in water emulsion may contain span 85 and/or lecithin and/or tricaprylin.

Typically for human administration QS21 and 3D-MPL will be present in a vaccine in the range of 1μg - 200μg, such as 10-100μg, preferably 10μg - 50μg per dose. Typically the

oil in water will comprise from 2 to 10% squalene, from 2 to 10% alpha tocopherol and from 0.3 to 3% tween 80. Preferably the ratio of squalene: alpha tocopherol is equal to or less than 1 as this provides a more stable emulsion. Span 85 may also be present at a level of 1%. In some cases it may be advantageous that the vaccines of the present invention will further contain a stabiliser.

Non-toxic oil in water emulsions preferably contain a non-toxic oil, e.g. squalane or squalene, an emulsifier, e.g. Tween 80, in an aqueous carrier. The aqueous carrier may be, for example, phosphate buffered saline.

A particularly potent adjuvant formulation involving QS21, 3D-MPL and tocopherol in an oil in water emulsion is described in WO 95/17210.

The present invention also provides a polyvalent vaccine composition comprising a vaccine formulation of the invention in combination with other antigens, in particular antigens useful for treating cancers, autoimmune diseases and related conditions. Such a polyvalent vaccine composition may include a TH-1 inducing adjuvant as hereinbefore described.

This invention also relates to the use of polynucleotides, in the form of primers derived from the polynucleotides of the present invention, and of polypeptides, in the form of antibodies or reagents specific for the polypeptide of the present invention, as diagnostic reagents.

The identification of genetic or biochemical markers in blood or tissues that will enable the detection of very early changes along the carcinogenesis pathway will help in determining the best treatment for the patient. Surrogate tumour markers, such as polynucleotide expression, can be used to diagnose different forms and states of cancer. The identification of expression levels of the polynucleotides of the invention will be useful in both the staging of the cancerous disorder and grading the nature of the cancerous tissue. The staging process monitors the advancement of the cancer and is determined on the presence or absence of malignant tissue in the areas biopsied. The polynucleotides of the invention can help to perfect the staging process by identifying markers for the aggresivity of a cancer, for example the presence in different areas of the body. The grading of the cancer describes how closely a tumour resembles normal tissue of its same type and is assessed by its cell

morphology and other markers of differentiation. The polynucleotides of the invention can be useful in determining the tumour grade as they can help in the determination of the differentiation status of the cells of a tumour.

The diagnostic assays offer a process for diagnosing or determining a susceptibility to cancers, autoimmune disease and related conditions through diagnosis by methods comprising determining from a sample derived from a subject an abnormally decreased or increased level of polypeptide or mRNA. This method of diagnosis is known as differential expression. The expression of a particular gene is compared between a diseased tissue and a normal tissue. A difference between the polynucleotide-related gene, mRNA, or protein in the two tissues is compared, for example in molecular weight, amino acid or nucleotide sequence, or relative abundance, indicates a change in the gene, or a gene which regulates it, in the tissue of the human that was suspected of being diseased.

Decreased or increased expression can be measured at the RNA level. PolyA RNA is first isolated from the two tissues and the detection of mRNA encoded by a gene corresponding to a differentially expressed polynucleotide of the invention can be detected by, for example, in situ hybridization in tissue sections, reverse trascriptase-PCR, using Northern blots containing poly A+ mRNA, or any other direct or inderect RNA detection method. An increased or decreased expression of a given RNA in a diseased tissue compared to a normal tissue suggests that the transcript and/or the expressed protein has a role in the disease. Thus detection of a higher or lower level of mRNA corresponding to SEQ ID NO 1 or 3 relative to normal level is indicative of the presence of cancer in the patient.

mRNA expression levels in a sample can be determined by generation of a library of expressed sequence tags (ESTs) from the sample. The relative representation of ESTs in the library can be used to assess the relative representation of the gene transcript in the starting sample. The EST analysis of the test can then be compared to the EST analysis of a reference sample to determine the relative expression levels of the polynucleotide of interest.

Other mRNA analyses can be carried out using serial analysis of gene expression (SAGE) methodology (Velculescu et. Al. Science (1995) 270:484) , differential display methodology (For example, US 5,776,683) or hybridization analysis which relies on the specificity of nucleotide interactions.

Alternatively, the comparison could be made at the protein level. The protein sizes in the two tissues may be compared using antibodies to detect polypeptides in Western blots of protein extracts from the two tissues. Expression levels and subcellular localization may also be detected immunologically using antibodies to the corresponding protein. Further assay techniques that can be used to determine levels of a protein, such as a polypeptide of the present invention, in a sample derived from a host are well-known to those of skill in the art. A raised or decreased level of polypeptide expression in the diseased tissue compared with the same protein expression level in the normal tissue indicates that the expressed protein may be involved in the disease.

In the assays of the present invention, the diagnosis can be determined by detection of gene product expression levels encoded by at least one sequence set forth in SEQ ID NOS: 1 or 3. A comparison of the mRNA or protein levels in a diseased versus normal tissue may also be used to follow the progression or remission of a disease.

A large number of polynucleotide sequences in a sample can be assayed using polynucleotide arrays. These can be used to examine differential expression of genes and to determine gene function. For example, arrays of the polynucleotide sequences SEQ ID NO: 1 or 3 can be used to determine if any of the polynucleotides are differentially expressed between a normal and cancer cell. In one embodiment of the invention, an array of oligonucleotides probes comprising the SEQ ID NO: 1 or 3 nucleotide sequence or fragments thereof can be constructed to conduct efficient screening of e.g., genetic mutations. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability (see for example: M.Chee et al., Science, Vol 274, pp 610-613 (1996)).

"Diagnosis" as used herein includes determination of a subject's susceptibility to a disease, determination as to whether a subject presently has the disease, and also the prognosis of a subject affected by the disease.

The present invention, further relates to a diagnostic kit for performing a diagnostic assay which comprises:

(a) a polynucleotide of the present invention, preferably the nucleotide sequence of SEQ ID NO: 1 or 3, or a fragment thereof;

(b) a nucleotide sequence complementary to that of (a);

(c) a polypeptide of the present invention, preferably the polypeptide of SEQ ID NO: 2 or 4, or a fragment thereof; or

(d) an antibody to a polypeptide of the present invention, preferably to the polypeptide of SEQ ID NO:2 or 4.

The nucleotide sequences of the present invention are also valuable for chromosomal localisation. The sequence is specifically targeted to, and can hybridize with, a particular location on an individual human chromosome. The mapping of relevant sequences to chromosomes according to the present invention is an important first step in correlating those sequences with gene associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found in, for example, V. McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).The differences in the cDNA or genomic sequence between affected and unaffected individuals can also be determined.

The polypeptides of the invention or their fragments or analogs thereof, or cells expressing them, can also be used as immunogens to produce antibodies immunospecific for polypeptides of the present invention. The term "immunospecific" means that the antibodies have substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art.

In a further aspect the invention provides an antibody immunospecific for a polypeptide according to the invention or an immunological fragment thereof as hereinbefore defined. Preferably the antibody is a monoclonal antibody

Antibodies generated against polypeptides of the present invention may be obtained by administering the polypeptides or epitope-bearing fragments, analogs or cells to an animal, preferably a non-human animal, using routine protocols. For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler, G. and Milstein, C., Nature (1975) 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, Immunology Today (1983) 4:72) and the EBV-hybridoma technique (Cole *et al.*, Monoclonal Antibodies and Cancer Therapy, 77-96, Alan R. Liss, Inc., 1985).

Techniques for the production of single chain antibodies, such as those described in U.S. Patent No. 4,946,778, can also be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms, including other mammals, may be used to express humanized antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or to purify the polypeptides by affinity chromatography.
The antibody of the invention may also be employed to prevent or treat cancer, particularly ovarian and colon cancer, autoimmune disease and related conditions.

Another aspect of the invention relates to a method for inducing or modulating an immunological response in a mammal which comprises inoculating the mammal with a polypeptide of the present invention, adequate to produce antibody and/or T cell immune response to protect or ameliorate the symptoms or progression of the disease. Yet another aspect of the invention relates to a method of inducing or modulating immunological response in a mammal which comprises, delivering a polypeptide of the present invention *via* a vector directing expression of the polynucleotide and coding for the polypeptide *in vivo* in order to induce such an immunological response to produce antibody to protect said animal from diseases.

It will be appreciated that the present invention therefore provides a method of treating abnormal conditions such as, for instance, cancer and autoimmune diseases, in particular, ovarian and colon cancer, related to either a presence of, an excess of, or an under-expression of, CASB619 polypeptide activity.

The present invention further provides for a method of screening compounds to identify those which stimulate or which inhibit the function of the CASB619 polypeptide. In general, agonists or antagonists may be employed for therapeutic and prophylactic purposes for such diseases as hereinbefore mentioned. Compounds may be identified from a variety of sources, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. Such agonists, antagonists or inhibitors so-identified may be natural or modified substrates, ligands, receptors, enzymes, etc., as the case may be, of the polypeptide; or may be structural or functional mimetics thereof (see Coligan *et al.*, Current Protocols in Immunology 1(2):Chapter 5 (1991)). Screening methods will be known to those skilled in the art. Further screening methods may be found in for example D. Bennett *et al.*, J Mol Recognition, 8:52-58 (1995); and K. Johanson *et al.*, J Biol Chem, 270(16):9459-9471 (1995) and references therein.

Thus the invention provides a method for screening to identify compounds which stimulate or which inhibit the function of the polypeptide of the invention which comprises a method selected from the group consisting of:

(a) measuring the binding of a candidate compound to the polypeptide (or to the cells or membranes bearing the polypeptide) or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound;

(b) measuring the binding of a candidate compound to the polypeptide (or to the cells or membranes bearing the polypeptide) or a fusion protein thereof in the presense of a labeled competitior;

(c) testing whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells or cell membranes bearing the polypeptide;

(d) mixing a candidate compound with a solution containing a polypeptide of claim 1, to form a mixture, measuring activity of the polypeptide in the mixture, and comparing the activity of the mixture to a standard; or

(e) detecting the effect of a candidate compound on the production of mRNA encoding said polypeptide and said polypeptide in cells, using for instance, an ELISA assay.

The polypeptide of the invention may be used to identify membrane bound or soluble receptors, if any, through standard receptor binding techniques known in the art. Well known screening methods may also be used to identify agonists and antagonists of the polypeptide of the invention which compete with the binding of the polypeptide of the invention to its receptors, if any.

Thus, in another aspect, the present invention relates to a screening kit for identifying agonists, antagonists, ligands, receptors, substrates, enzymes, etc. for polypeptides of the present invention; or compounds which decrease or enhance the production of such polypeptides, which comprises:

(a) a polypeptide of the present invention;

(b) a recombinant cell expressing a polypeptide of the present invention;

(c) a cell membrane expressing a polypeptide of the present invention; or

(d) antibody to a polypeptide of the present invention;

which polypeptide is preferably that of SEQ ID NO:2.

It will be readily appreciated by the skilled artisan that a polypeptide of the present invention may also be used in a method for the structure-based design of an agonist, antagonist or inhibitor of the polypeptide, by:

(a) determining in the first instance the three-dimensional structure of the polypeptide;

(b) deducing the three-dimensional structure for the likely reactive or binding site(s) of an agonist, antagonist or inhibitor;

(c) synthesing candidate compounds that are predicted to bind to or react with the deduced binding or reactive site; and

(d) testing whether the candidate compounds are indeed agonists, antagonists or inhibitors.

Gene therapy may also be employed to effect the endogenous production of CASB619 polypeptide by the relevant cells in the subject. For an overview of gene therapy, see Chapter 20, Gene Therapy and other Molecular Genetic-based Therapeutic Approaches, (and references cited therein) in Human Molecular Genetics, T Strachan and A P Read, BIOS Scientific Publishers Ltd (1996).

Vaccine preparation is generally described in Pharmaceutical Biotechnology, Vol.61 Vaccine Design - the subunit and adjuvant approach, edited by Powell and Newman, Plenum Press, 1995. New Trends and Developments in Vaccines, edited by Voller et al., University Park Press, Baltimore, Maryland, U.S.A. 1978. Encapsulation within liposomes is described, for example, by Fullerton, U.S. Patent 4,235,877. Conjugation of proteins to macromolecules is disclosed, for example, by Likhite, U.S. Patent 4,372,945 and by Armor et al., U.S. Patent 4,474,757.

The amount of protein in each vaccine dose is selected as an amount which induces an immunoprotective response without significant, adverse side effects in typical vaccinees. Such amount will vary depending upon which specific immunogen is employed. Generally, it is expected that each dose will comprise 1-1000μg of protein, preferably 2-100μg, most preferably 4-40μg. An optimal amount for a particular vaccine can be ascertained by standard studies involving observation of antibody titres and other responses in subjects. Following an initial vaccination, subjects may receive a boost in about 4 weeks.

"Isolated" means altered "by the hand of man" from the natural state. If an "isolated" composition or substance occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living animal is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein.

"Polynucleotide" generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA including single and double stranded regions.

"Variant" refers to a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis.

"Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in (Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., SIAM *J. Applied Math.*, *48:* 1073 (1988). Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer

program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG program package (Devereux, J., et al., Nucleic Acids Research *12(1):* 387 (1984)), BLASTP, BLASTN, and FASTA (Atschul, S.F. et al., J. Molec. Biol. *215:* 403-410 (1990). The BLAST X program is publicly available from NCBI and other sources (BLAST Manual, Altschul, S., *et al.*, NCBI NLM NIH Bethesda, MD 20894; Altschul, S., *et al.*, J. Mol. Biol. *215*: 403-410 (1990). The well known Smith Waterman algorithm may also be used to determine identity.

The preferred algorithm used is FASTA. The preferred parameters for polypeptide or polynuleotide sequence comparison using this algorithm include the following:
Gap Penalty:12
Gap extension penalty: 4
Word size: 2, max 6

Preferred parameters for polypeptide sequence comparison with other methods include the following:
1) Algorithm: Needleman and Wunsch, J. Mol Biol. 48: 443-453 (1970)
Comparison matrix: BLOSSUM62 from Hentikoff and Hentikoff, Proc. Natl. Acad. Sci. USA. 89:10915-10919 (1992)
Gap Penalty: 12
Gap Length Penalty: 4

A program useful with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison WI. The aforementioned parameters are the default parameters for polypeptide comparisons (along with no penalty for end gaps).

Preferred parameters for polynucleotide comparison include the following:
1) Algorithm: Needleman and Wunsch, J. Mol Biol. 48: 443-453 (1970)
Comparison matrix: matches = +10, mismatch = 0
Gap Penalty: 50
Gap Length Penalty: 3

A program useful with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison WI. The aforementioned parameters are the default parameters for polynucleotide comparisons.

By way of example, a polynucleotide sequence of the present invention may be identical to the reference sequence of SEQ ID NO:1, that is be 100% identical, or it may include up to a certain integer number of nucleotide alterations as compared to the reference sequence. Such alterations are selected from the group consisting of at least one nucleotide deletion, substitution, including transition and transversion, or insertion, and wherein said alterations may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence. The number of nucleotide alterations is determined by multiplying the total number of nucleotides in SEQ ID NO:1 by the numerical percent of the respective percent identity(divided by 100) and subtracting that product from said total number of nucleotides in SEQ ID NO:1, or:

$$n_n \leq x_n - (x_n \bullet y),$$

wherein n_n is the number of nucleotide alterations, x_n is the total number of nucleotides in SEQ ID NO:1, and y is, for instance, 0.70 for 70%, 0.80 for 80%, 0.85 for 85%, 0.90 for 90%, 0.95 for 95%,etc., and wherein any non-integer product of x_n and y is rounded down to the nearest integer prior to subtracting it from x_n. Alterations of a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2 may create nonsense, missense or frameshift mutations in this coding sequence and thereby alter the polypeptide encoded by the polynucleotide following such alterations.

Similarly, a polypeptide sequence of the present invention may be identical to the reference sequence of SEQ ID NO:2, that is be 100% identical, or it may include up to a certain integer number of amino acid alterations as compared to the reference sequence such that the % identity is less than 100%. Such alterations are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion, and wherein said alterations may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere

between those terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence. The number of amino acid alterations for a given % identity is determined by multiplying the total number of amino acids in SEQ ID NO:2 by the numerical percent of the respective percent identity(divided by 100) and then subtracting that product from said total number of amino acids in SEQ ID NO:2, or:

$$n_a \leq x_a - (x_a \bullet y),$$

wherein n_a is the number of amino acid alterations, x_a is the total number of amino acids in SEQ ID NO:2, and y is, for instance 0.70 for 70%, 0.80 for 80%, 0.85 for 85% etc., and wherein any non-integer product of x_a and y is rounded down to the nearest integer prior to subtracting it from x_a.

"Homolog" is a generic term used in the art to indicate a polynucleotide or polypeptide sequence possessing a high degree of sequence relatedness to a subject sequence. Such relatedness may be quntified by determining the degree of identity and/or similarity between the sequences being compared as hereinbefore described. Falling within this generic term are the terms "ortholog", meaning a polynucleotide or polypeptide that is the functional equivalent of a polynucleotide or polypeptide in another species and "paralog" meaning a functionally similar sequence when considered within the same species.

Examples

Example 1

Real-time RT-PCR analysis

Real-time RT-PCR (U. Gibson. 1996. Genome Research: 6,996) is used to compare mRNA transcript abundance of the candidate antigen in matched tumour and normal colon tissues from multiple patients. In addition, mRNA levels of the candidate gene in a panel of normal tissues are evaluated by this approach.

Total RNA from normal and tumour colon is extracted from snap frozen biopsies using TriPure reagent (Boehringer). Total RNA from normal tissues is purchased from

InVitrogen or is extracted from snap frozen biopsies using TriPure reagent. Poly-A+ mRNA is purified from total RNA after DNAase treatment using oligo-dT magnetic beads (Dynal). Quantification of the mRNA is performed by spectrofluorimetry (VersaFluor, BioRad) using SybrII dye (Molecular Probes). Primers for real-time PCR amplification are designed with the Perkin-Elmer Primer Express software using default options for TaqMan amplification conditions.

Real-time reactions are assembled according to standard PCR protocols using 2 ng of purified mRNA for each reaction. SybrI dye (Molecular Probes) is added at a final dilution of 1/75000 for real-time detection. Amplification (40 cycles) and real-time detection is performed in a Perkin-Elmer Biosystems PE7700 system using conventional instrument settings. Ct values are calculated using the PE7700 Sequence Detector software. Two Ct values are obtained for each patient sample: the tumour Ct (CtT) and the matched normal colon Ct (CtN). Ct values obtained by real-time PCR are log-linearly related to the copy number of the target template. As the efficiency of PCR amplification under the prevailing experimental conditions is close to the theoretical amplification efficiency, 2 (CtN-CtT) is an estimate of the relative transcript levels in the two tissues (i.e. fold mRNA over-expression in tumor). Real-time PCR reactions are performed on biopsies from 24 patients. The level of mRNA over-expression is calculated as described for each patient. average level of mRNA over-expression for the candidate antigen and the proportion of patients over-expressing the candidate antigen is then calculated from this data set. The individual values are standardised with respect to actin in the same sample (ratio), as seen in Figure 1. A value of 1 thus corresponds to the same level of actin expression. The results are shown in a logarithmic scale.

A series of 48 normal tissue samples, representing 28 different tissues, were also tested by the same procedure. Ct values for the candidate antigen were compared to those of actin obtained with the same tissue sample. The results, standardized with respect to actin, are shown in Figure 2.

Real-time PCR results in colon cancer/normal colon sample

Summary

Patients over-expressing CASB619	Average level of over-expression

in colon tumours (%)	in colon tumours (fold)
16/24 (67%)	208

Conclusion: CASB619 is overexpressed in 67% of colon cancer samples with respect to the adjacent normal colon, at an average rate of nearly 200 fold. The expression in normal tissues is restricted to other digestive tract tissues, as well as mainly trachea and testis.

Example 2.

DNA microarrays

DNA micro-arrays are used to examine mRNA expression profiles of large collections of genes in multiple samples. This information is used to complement the data obtained by real-time PCR and provides an independent measure of gene expression levels in tumors and normal tissues.

Examples of current technologies for production of DNA micro-arrays include 1) The Affymetrix "GeneChip" arrays in which oligonucleotides are synthetized on the surface of the chip by solid phase chemical synthesis using a photolithographic process 2) DNA spotting technology in which small volumes of a DNA solution are robotically deposited and then immobilized onto the surface of a solid phase (e.g. glass). In both instances, the chips are hybridized with cDNA or cRNA which has been extracted from the tissue of interest (e.g. normal tissue, tumour etc...) and labeled with radioactivity or with a fluorescent reporter molecule. The labeled material is hybridized to the chip and the amount of probe bound to each sequence on the chip is determined using a specialized scanner. The experiment can be set-up with a single fluorescent reporter (or radioactivity) or, alternatively, can be performed using two fluorescent reporters. In this latter case, each of the two samples is labeled with one of the reporter molecules. The two labeled samples are then hybridized competitively to the sequences on the DNA chip. The ratio of the two fluorescent signals is determined for each sequence on the chip. This ratio is used to calculate the relative abundance of the transcript in the two samples. Detailed protocols are available from a number of sources including "DNA Microarrays: A practical approach. Schena M. Oxford University Press 1999" and the World Wide Web

(http://cmgm.stanford.edu/pbrown/protocols/index.html), http://arrayit.com/DNA-Microarray-Protocols/) and specialized distributors (e.g. Affymetrix).

Example 3.

EST profiles

A complementary approach to experimental antigen tissue expression characterization is to explore the human "Expressed Sequence Tags" (ESTs) database. ESTs are small fragments of cDNA made from a collection of mRNA extracted from a particular tissue or cell line. Such database currently provide a massive amount of ESTs (106) from several hundreds of cDNA tissue libraries, including tumoral tissues from various types and states of disease. By means of informatics tools (Blast), a comparison search of the CASB616 sequence is performed in order to have further insight into tissue expression.

EST distribution of CASB619

DbEST accession	ATG lib ID	Description
NCBI:1202616	937	NCI_CGAP_Co2 (colon vilous adenoma)
NCBI:1202659	937	NCI_CGAP_Co2 (colon vilous adenoma)
NCBI:1208269	935	NCI_CGAP_Pr4.1 (prostatic intraepithelial
NCBI:1152744	888	NCI_CGAP_Pr6 (prostatic intraepithelial neoplasia)
NCBI:1157532	910	NCI_CGAP_Pr22 (normal prostate)
NCBI:1178873	882	NCI_CGAP_Co3 (12 pooled tumor colon)
NCBI:1040601	628	testis NHT
NCBI:1056221	628	testis NHT
NCBI:1298131	910	NCI_CGAP_Pr22 (normal prostate)
NCBI:976517	715	epididyme
NCBI:1618753	417	parathyroid tumor
NCBI:1737305	895	NCI_CGAP_Br2 (pooled breast tumor tissues)
NCBI:1738600	895	NCI_CGAP_Br2 (pooled breast tumor tissues)
NCBI:1767617	628	testis NHT
NCBI:1889992	424	fetal heart
NCBI:1907671	628	testis NHT
NCBI:2074937	1076	NCI_CGAP_Lu5 View stats
NCBI:2147596	1461	NCI_CGAP_Ut1
NCBI:2147632	1461	NCI_CGAP_Ut1
NCBI:2308815	1462	NCI_CGAP_Ut2
NCBI:2441896	1410	NCI_CGAP_Pr28
NCBI:2447648	1463	NCI_CGAP_Ut3
NCBI:2381271	1461	NCI_CGAP_Ut1
NCBI:2582781	1461	NCI_CGAP_Ut1
NCBI:2583713	1461	NCI_CGAP_Ut1
NCBI:2585593	1410	NCI_CGAP_Pr28

NCBI:2587322	1463	NCI_CGAP_Ut3
NCBI:2601136	1461	NCI_CGAP_Ut1
NCBI:2824305	1463	NCI_CGAP_Ut3
NCBI:3007221	1461	NCI_CGAP_Ut1
NCBI:3078572	2301	NCI CGAP Pit1
NCBI:3086989	1461	NCI_CGAP_Ut1
NCBI:3087235	1461	NCI_CGAP_Ut1
NCBI:3045696	1661	NCI_CGAP_Lu19
NCBI:3218161	2467	NCI CGAP Co20
NCBI:3291383	2508	NCI_CGAP_Sub3
NCBI:3028913	2107	BT130
NCBI:3655706	1447	NCI_CGAP_Co14
NCBI:2553046	1728	Soares_Dieckgraefe_colon_NHCD
NCBI:979345	781	endometrial tumor
NCBI:3289738	2508	NCI_CGAP_Sub3
NCBI:614598	464	?
NCBI:978076	781	endometrial tumor
NCBI:979060	781	endometrial tumor
NCBI:1616799	417	parathyroid tumor
NCBI:978445	781	endometrial tumor

In summary, 93% of the ESTs matching to CASB619 originate either from tumor tissues, fetal tissues, or normal reproductive organs. This suggests that this gene is more frequently expressed in these tissues.

Example 4

Northern-Southern blot analysis

Limited amounts of mixed tumour and matched normal colon cDNA are amplified by Advantage PCR (see above). Messenger RNA from multiple normal tissues is also amplified using the same procedure. The amplified cDNA (1 μg) is electrophoresed on a 1.2% agarose gel and transferred onto a nylon membrane. The membrane is hybridised (AlkPhos Direct System) with a probe prepared using a fragment of the candidate TAA cDNA. Northern-Southern analysis provides information on transcript size, presence of splice variants and transcript abundance in tumour and normal tissues.

Example 5

Northern Blot Analysis

Northern blots are produced according to standard protocols using 1 μg of poly A+ mRNA. Radioactive probes are prepared using the Ready-to-Go system (Pharmacia).

Example 6

Identification f the full length cDNA sequence

Colon tumour cDNA libraries are constructed using the Lambda Zap II system (Stratagene) from 5 µg of polyA+ mRNA. The supplied protocol is followed except that SuperscriptII (Life Technologies) is used for the reverse transcription step. Oligo dT-primed and random-primed libraries are constructed. About 1.5 x106 independent phages are plated for each screening of the library. Phage plaques are transferred onto nylon filters and hybridised using a cDNA probe labelled with AlkPhos Direct. Positive phages are detected by chemiluminescence. Positive phage are excised from the agar plat, eluted in 500µl SM buffer and confirmed by gene-specific PCR. Eluted phages are converted to single strand M13 bacteriophage by in vivo excision. The bacteriophage is then converted to double strand plasmid DNA by infection of E. coli. Infected bacteria are plated and submitted to a second round of screening with the cDNA probe. Plasmid DNA is purified from positive bacterial clones and sequenced on both strands.

When the full length gene cannot be obtained directly from the cDNA library, missing sequence is isolated using RACE technology (Marathon Kit, ClonTech.). This approach relies on reverse transcribing mRNA into double strand cDNA, ligating linkers onto the ends of the cDNA and amplifying the desired extremity of the cDNA using a gene-specific primer and one of the linker oligonucleotides. Marathon PCR products are cloned into a plasmid (pCRII-TOPO, InVitrogen) and sequenced.

The obtained sequence (SEQ ID NO:1) has a putative open reading frame of 1013 amino acids (SEQ ID NO:2). The deduced protein sequence was submitted to prediction algorithms for cellular localisation (PSORT: http://psort.nibb.ac.jp/ and TopPred: http://www.biokemi.su.se/~server/toppred2/toppred_source.html). CASB619 seems to have a peptide signal, and one to three transmembrane domains (low confidence prediction). No other motif or domain was found.

Example 7 :

7.1 Expression and purification of tumour-specific antigens

Expression in microbial hosts, or alternatively in vitro transcription/translation, is used to produce the antigen of the invention for vaccine purposes and to produce protein fragments or whole protein for rapid purification and generation of antibodies needed for

characterization of the naturally expressed protein by immunohistochemistry or for follow-up of purification.

Recombinant proteins may be expressed in two microbial hosts, E. coli and in yeast (such as Saccharomyces cerevisiae or Pichia pastoris).Pichia. This allows the selection of the expression system with the best features for this particular antigen production. In general, the recombinant antigen will be expressed in E. coli and the reagent protein expressed in yeast.

The expression strategy first involves the design of the primary structure of the recombinant antigen. In general an expression fusion partner (EFP) is placed at the N terminal extremity to improve levels of expression that could also include a region useful for modulating the immunogenic properties of the antigen, an immune fusion partner (IFP). In addition, an affinity fusion partner (AFP) useful for facilitating further purification is included at the C-terminal end.
When the recombinant strains are available, the recombinant product is characterized by the evaluation of the level of expression and the prediction of further solubility of the protein by analysis of the behavior in the crude extract.

After growth on appropriate culture medium and induction of the recombinant protein expression, total extracts are analyzed by SDS-PAGE. The recombinant proteins are visualized in stained gels and identified by Western blot analysis using specific antibodies.

A comparative evaluation of the different versions of the expressed antigen will allow the selection of the most promising candidate that is to be used for further purification and immunological evaluation.

The purification scheme follows a classical approach based on the presence of an His affinity tail in the recombinant protein. In a typical experiment the disrupted cells are filtered and the acellular extracts loaded onto an Ion Metal Affinity Chromatography (IMAC; Ni++NTA from Qiagen) that will specifically retain the recombinant protein.

The retained proteins are eluted by 0-500 mM Imidazole gradient (possibly in presence of a detergent) in a phosphate buffer.

7.2 Antibody production and immunohistochemistry

Small amounts of relatively purified protein can be used to generate immunological tools in order to

a) detect the expression by immunohistochemistry in normal or cancer tissue sections;

b) detect the expression, and to follow the protein during the purification process (ELISA/ Western Blot); or

c) characterise/ quantify the purified protein (ELISA).

7.2.1 Polyclonal antibodies:

Immunization

2- 3 Rabbits are immunised , intramuscularly (I.M.) , 3 times at 3 weeks intervals with 100µg of protein, formulated in the adjuvant 3D-MPL/QS21. Three weeks after each immunisation a blood sample is taken and the antibody titer estimated in the serum by ELISA using the protein as coating antigen following a standard protocol.

ELISA

96 well microplates (maxisorb Nunc) are coated with 5µg of protein overnight at 4°C. After 1hour saturation at 37°C with PBS NCS 1%, serial dilution of the rabbit sera is added for 1H 30 at 37°C (starting at 1/10). After 3 washings in PBS Tween, anti rabbit biotinylated anti serum (Amersham) is added (1/5000). Plates are washed and peroxydase coupled streptavidin (1/5000) is added for 30 min at 37°C. After washing, 50µl TMB (BioRad) is added for 7 min and the reaction then stopped with H2SO4 0.2M. The OD can be measured at 450 nm and midpoint dilutions calculated by SoftmaxPro.

7.2.2 Monoclonal antibodies:

Immunization

5 BALB/c mice are immunized 3 times at 3 week intervals with 5 µg of purified protein. Bleedings are performed 14 days post II and 1 week post 3. The sera are tested by Elisa on purified protein used as coated antigen. Based on these results (midpoint dilution > 10000) one mouse is selected for fusion

Fusion/ HATselection

Spleen cells are fused with the SP2/0 myeloma according to a standard protocol using PEG 40% and DMSO 5%. Cells are then seeded in 96 well plates 2.5 x104 - 105 cells/well and resistant clones will be selected in HAT medium. The supernatant of these hybridomas will be tested for their content in specific antibodies and when positive, will be submitted to 2 cycles of limited dilution . After 2 rounds of screening, 3 hybridomas will be chosen for ascitis production.

7.2.3 Immunohistochemistry

When antibodies are available, immuno staining is performed on normal or cancer tissue sections, in order to determine :

- ◊ the level of expression of the antigen of the invention in cancer relative to normal tissue or
- ◊ the proportion of cancer of a certain type expressing the antigen
- ◊ if other cancer types also express the antigen
- ◊ the proportion of cells expressing the antigen in a cancer tissue

Tissue sample preparation

After dissection, the tissue sample is mounted on a cork disk in OCT compound and rapidly frozen in isopentane previously super cooled in liquid nitrogen (-160°C). The block will then be conserved at -70°C until use. 7-10µm sections will be realised in a cryostat chamber (-20, -30°C).

Staining

Tissue sections are dried for 5 min at room Temperature (RT), fixed in acetone for 10min at RT, dried again, and saturated with PBS 0.5% BSA 5% serum. After 30 min at RT either a direct or indirect staining is performed using antigen specific antibodies. A direct staining leads to a better specificity but a less intense staining whilst an indirect staining leads to a more intense but less specific staining.

7.3 Analysis of human cellular immune resp nses to the antigen of the invention

The immunological relevance of the antigen of the invention can be assessed by in vitro priming of human T cells. All T cell lymphocyte lines and dendritic cells are derived from PBMCs (peripheral blood mononuclear cells) of healthy donors (preferred HLA-A2 subtype). An HLA-A2.1/Kb transgenic mice is also used for screening of HLA-A2.1 peptides.

Newly discovered antigen-specific CD8+ T cell lines are raised and maintained by weekly in vitro stimulation. The lytic activity and the □-IFN production of the CD8 lines in response to the antigen or antigen derived-peptides is tested using standard procedures.

Two strategies to raise the CD8+ T cell lines are used: a peptide-based approach and a whole gene-based approach. Both approaches require the full-length cDNA of the newly discovered antigen in the correct reading frame to be either cloned in an appropriate delivery system or to be used to predict the sequence of HLA binding peptides.

Peptide-based approach

The HLA-A2 binding peptide sequences are predicted either by the Parker's algorithm (Parker, K. C., M. A. Bednarek, and J. E. Coligan. 1994. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. J. Immunol. 152:163 and http://bimas.dcrt.nih.gov/molbio/hla_bind/) or the Rammensee method (Rammensee, Friede, Stevanovic, MHC ligands and peptide motifs: 1st listing, Immunogenetics 41, 178-228, 1995 ; Rammensee, Bachmann, Stevanovic: MHC ligands and peptide motifs. Landes Bioscience 1997, and http://134.2.96.221/scripts/hlaserver.dll/home.htm). Peptides are then screened in the HLA-A2.1/Kb transgenic mice model (Vitiello et al.).

a) Predicted epitopes binding the HLA_A0201 allele :

a.1) HLA-A*0201 nonamers

Position	1	2	3	4	5	6	7	8	9	Rammense e score	Parker score°	SEQ ID NO
848	F	L	W	E	S	A	A	A	C		777.681	68
24	R	L	W	R	L	L	L	W	A		521.615	5
761	S	L	A	D	R	L	I	G	V	30	655.875	6

893	S	L	P	E	Q	R	V	T	I	26		7
886	K	L	C	S	G	G	I	S	L	25		8
853	A	A	A	C	P	L	C	S	V	25		9
674	S	A	L	A	N	T	V	T	L	24		10
499	F	V	F	E	T	L	C	S	V	24	976.762	11
129	E	L	P	H	G	F	A	S	L	23		12
973	L	I	F	T	S	K	K	S	L	22		13
936	K	L	E	Y	K	Y	S	K	L	22		14
903	K	T	I	D	F	W	L	K	V	22		15
860	S	V	A	D	Y	H	A	I	V	22		16
830	L	L	L	P	G	T	C	S	D	22		17
675	A	L	A	N	T	V	T	L	A	22		18
503	T	L	C	S	V	N	C	E	L	22		19
169	N	T	D	E	C	T	A	T	L	22		20
81	S	L	P	D	P	V	K	G	T	22		21
980	S	L	F	G	K	I	K	S	F	21		22
918	C	T	A	I	L	L	T	V	L	21		23
867	I	V	S	S	C	V	A	G	I	21		24
710	K	M	S	V	C	T	D	N	V	21		25
259	V	L	V	R	N	I	A	I	T	21		26
234	E	L	N	R	G	N	N	V	L	21		27
175	A	T	L	M	Y	A	V	N	L	21		28
70	V	A	V	P	H	T	P	G	L	21		29
24	R	L	W	R	L	L	L	W	A	21		30
914	S	A	G	T	C	T	A	I	L	20		31
891	G	I	S	L	P	E	Q	R	V	20		32
824	K	T	V	P	G	S	L	L	L	20		33
765	R	L	I	G	V	T	T	D	M	20		34
681	T	L	A	G	G	P	S	F	T	20		35
539	Y	I	I	E	E	N	T	T	T	20		36
264	I	A	I	T	G	V	A	Y	T	20		37
38	V	T	Q	G	T	G	P	E	L	20		38

° Estimate of half time of disassociation of a molecule containing this subsequence

a.2) HLA A02_01 decamers

Position	1	2	3	4	5	6	7	8	9	10	Rammen see score	Parker score°	SEQ ID NO
980	S	L	F	G	K	I	K	S	F	T		151.648	39
866	A	I	V	S	S	C	V	A	G	I	25		40
852	S	A	A	A	C	P	L	C	S	V	24		41
786	H	L	E	S	L	G	I	P	D	V	24		42
571	K	I	Y	S	I	N	V	T	N	V	24	246.353	43
761	S	L	A	D	R	L	I	G	V	T	23		44
626	I	L	K	A	H	Q	P	Y	G	V	23		5
485	V	M	A	D	T	E	N	K	E	V	23	350.117	46
29	L	L	W	A	G	T	A	F	Q	V	23	5691.997	47
916	G	T	C	T	A	I	L	L	T	V	22		48
778	I	T	S	P	A	E	L	F	H	L	22		49
766	L	I	G	V	T	T	D	M	T	L	22		50
428	T	L	P	T	N	M	E	T	T	V	22		51

350	L	M	Y	K	W	A	K	P	K	I	22		52
972	D	L	I	F	T	S	K	K	S	L	21		53
692	G	L	K	Y	F	H	H	F	T	L	21		54
644	G	T	K	N	N	K	I	H	S	L	21		55
465	S	D	N	D	F	M	I	L	T	L	21		56
260	L	V	R	N	I	A	I	T	G	V	21		57
77	G	L	C	T	S	L	P	D	P	V	21		58
949	T	L	K	D	C	D	L	P	A	A	20		59
795	V	I	F	F	Y	R	S	N	D	V	20		60
733	S	I	T	A	Y	V	C	Q	A	V	20		61
712	S	V	C	T	D	N	V	T	D	L	20		62
702	S	L	C	G	N	Q	G	R	K	M	20		63
432	N	M	E	T	T	V	L	S	G	I	20		64
263	N	I	A	I	T	G	V	A	Y	T	20		65
121	G	I	R	F	D	E	W	D	E	L	20		66

° Estimate of half time of disassociation of a molecule containing this subsequence

HLA_A0205

Position	1	2	3	4	5	6	7	8	9	Parker score°	SEQ ID NO
499	F	V	F	E	T	L	C	S	V	216	11

HLA A0203

Position	1	2	3	4	5	6	7	8	9	10	Rammen see score	SEQ ID NO
846	F	H	F	L	W	E	S	A	A	A	27	67

Briefly, transgenic mice are immunized with adjuvanted HLA-A2 peptides, those unable to induce a CD8 response (as defined by an efficient lysis of peptide-pulsed autologous spleen cells) will be further analyzed in the human system.

Human dendritic cells (cultured according to Romani et al.) will be pulsed with peptides and used to stimulate CD8-sorted T cells (by Facs). After several weekly stimulations, the CD8 lines will be first tested on peptide-pulsed autologous BLCL (EBV-B transformed cell lines). To verify the proper in vivo processing of the peptide, the CD8 lines will be tested on cDNA-transfected tumour cells (HLA-A2 transfected LnCaP, Skov3 or CAMA tumour cells).

Whole gene-based approach

CD8+ T cell lines will be primed and stimulated with either gene-gun transfected dendritic cells, retrovirally transduced B7.1-transfected fibroblasts, recombinant pox

virus (Kim et al.) or adenovirus (Butterfield et al.) infected dendritic cells. Virus infected cells are very efficient to present antigenic peptides since the antigen is expressed at high level but can only be used once to avoid the over-growth of viral T cells lines.

After alternated stimulations, the CD8+ lines are tested on cDNA-transfected tumour cells as indicated above. Peptide specificity and identity is determined to confirm the immunological validation.

References

Vitiello et al. (L. Sherman), J. Exp. Med., J. Exp. Med, 1991, 173:1007-1015.

Romani et al., J. Exp. Med., 1994, 180:83-93.

Kim et al., J. Immunother., 1997, 20:276-286.

Butterfield et al., J. Immunol., 1998, 161:5607-5613.

All publications, including but not limited to patents and patent applications, cited in this specification are herein incorporated by reference as if each individual publication were specifically and individually indicated to be incorporated by reference herein as though fully set forth.

Claims

1. An isolated polypeptide comprising an amino acid sequence which has at least 70% identity to the amino acid sequence of SEQ ID NO:2 over the entire length of of SEQ ID NO:2.

2. An isolated polypeptide as claimed in claim 1 in which the amino acid sequence has at least 95% identity to SEQ ID NO:2.

3. The polypeptide as claimed in claim 1 comprising the amino acid sequence of SEQ ID NO:2.

4. The isolated polypeptide of SEQ ID NO:2.

5. A polypeptide comprising an immunogenic fragment of a polypeptide as claimed in any one of claims 1 to 4 in which the immunogenic activity of the immunogenic fragment is substantially the same as the polypeptide of SEQ ID NO:2

6. A polypeptide as claimed in any of claims 1 to 5 wherein said polypeptide is part of a larger fusion protein.

7. A polypeptide as claimed in any of claims 1 to 6 chemically conjugated to a carrier protein.

8. An isolated polynucleotide encoding a polypeptide as claimed in any of claims 1 to 6.

9. An isolated polynucleotide comprising a nucleotide sequence encoding a polypeptide that has at least 70% identity to the amino acid sequence of SEQ ID NO:2, over the entire length of SEQ ID NO:2; or a nucleotide sequence complementary to said isolated polynucleotide.

10. An isolated polynucleotide comprising a nucleotide sequence that has at least 70% identity to a nucleotide sequence encoding a polypeptide of SEQ ID NO:2, over the entire coding region; or a nucleotide sequence complementary to said isolated polynucleotide.

11. An isolated polynucleotide which comprises a nucleotide sequence which has at least 70% identity to that of SEQ ID NO:1 over the entire length of SEQ ID NO:1; or a nucleotide sequence complementary to said isolated polynucleotide.

12. The isolated polynucleotide as defined in any one of claims 8 to 11 in which the identity is at least 95%.

13. An isolated polynucleotide selected from:

(a) a polynucleotide comprising a nucleotide sequence encoding the polypeptide of SEQ ID NO:2;

(b) the polynucleotide of SEQ ID NO:1; and

(c) a polynucleotide obtainable by screening an appropriate library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO:1 or a fragment thereof said polynucleotide encoding a protein which has similar immunogenic properties to those of the protein of sequence ID NO:2

or a nucleotide sequence complementary to said isolated polynucleotide

14. An expression vector or a recombinant live microorganism comprising an isolated polynucleotide according to any one of claims 8 - 13.

15. A host cell comprising the expression vector of claim 14 or the isolated polynucleotide of claims 8 to 13.

16. A process for producing a polypeptide of claims 1 to 7 comprising culturing a host cell of claim 15 under conditions sufficient for the production of said polypeptide and recovering the polypeptide from the culture medium.

17. A vaccine comprising an effective amount of the polypeptide of any one of claims 1 to 7 and a pharmaceutically acceptable carrier.

18. A vaccine comprising an effective amount of the polynucleotide of any one of claims 8 to 13 and a pharmaceutically effective carrier.

19. A vaccine comprising an effective amount of antigen presenting cells, modified by in vitro loading with a polypeptide of any one of claims 1 to 7, or genetically modified in vitro to express a polypeptide of claims 1 to 7 and a pharmaceutically effective carrier.

20. A vaccine as claimed in any one of claims 17 to 19 which additionally comprises a TH-1 inducing adjuvant.

21. A vaccine as claimed in claim 20 in which the TH-1 inducing adjuvant is selected from the group of adjuvants comprising: 3D-MPL, QS21, a mixture of QS21 and cholesterol, and a CpG oligonucleotide.

22. An antibody immunospecific for the polypeptide or immunological fragment as claimed in any one of claims 1 to 5.

23. A method for screening to identify compounds which stimulate or which inhibit the function of the polypeptide of any one of claims 1 to 5 which comprises a method selected from the group consisting of:

(a) measuring the binding of a candidate compound to the said polypeptide (or to the cells or membranes bearing the polypeptide) or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound;

(b) measuring the binding of a candidate compound to the said polypeptide (or to the cells or membranes bearing the polypeptide) or a fusion protein thereof in the presence of a labeled competitior;

(c) testing whether the candidate compound results in a signal generated by activation or inhibition of the said polypeptide, using detection systems appropriate to the cells or cell membranes bearing the polypeptide;

(d) mixing a candidate compound with a solution containing a polypeptide of any one of claims 1 to 7, to form a mixture, measuring activity of the polypeptide in the mixture, and comparing the activity of the mixture to a standard; or

(e) detecting the effect of a candidate compound on the production of mRNA encoding said polypeptide and said polypeptide in cells, using for instance, an ELISA assay.

24. A method for the treatment of a subject by immunoprophylaxis or therapy comprising *in vitro* induction of immune responses to a molecule of any one of claims 1 to 5, using *in vitro* incubation of the polypeptide of any one of claims 1 to 7 or the polynucleotide of any one of claims 8 to 13 with cells from the immune system of a mammal, and reinfusing these activated immune cells to the mammal for the treatment of disease.

25. A method as claimed in claim 24 wherein the treatment is for ovarian or colon cancer.

26. An agonist or antagonist to the polypeptide of claims 1 to 5.

27. A compound which is:
(a) an agonist or antagonist to the polypeptide of claims 1 to 5;
(b) isolated polynucleotide of claims 8 to 13; or
(c) a nucleic acid molecule that modulates the expression of the nucleotide sequence encoding the polypeptide of any one of claims 1 to 5;
for use in therapy.

28. A process for diagnosing a disease or a susceptibility to a disease in a subject related to expression or activity of a polypeptide of any one of claims 1 to 5 in a subject comprising analyzing for the presence or amount of said polypeptide in a sample derived from said subject.

29. A process for diagnosing a disease or a susceptibility to a disease in a subject related to expression or activity of a polynucleotide of any one of claims 8 to 13 in a subject comprising analyzing for the presence or amount of said polynucleotide in a sample derived from said subject.

30. A process for diagnosing the presence of colon cancer or a susceptibility to colon cancer in a subject related to expression or activity of a polypeptide of any one of claims 1 to 5 in a subject comprising analyzing for the presence or amount of said polypeptide in a sample derived from said subject.

31. A process for diagnosing the presence of colon cancer or a susceptibility to colon cancer in a subject related to expression or activity of a polynucleotide of any one of claims 8 to 13 in a subject comprising analyzing for the presence or amount of said polynucleotide in a sample derived from said subject.

32. An isolated polynucleotide selected from the group consisting of:
(a) an isolated polynucleotide comprising a nucleotide sequence which has at least 70% identity to SEQ ID NO:3 over the entire length of SEQ ID NO:3;
(b) an isolated polynucleotide comprising the polynucleotide of SEQ ID NO:3;
(c) the polynucleotide of SEQ ID NO:3.

33. A live vaccine composition comprising an expression vector or recombinant live micro-organism according to claim 14.

34. Use of a polynucleotide as claimed in any one of claims 8 to 13 for the manufacture of a medicament in the treatment of carcinoma.

35. Use of a polynucleotide as claimed in any one of claims 8 to 13 for the manufacture of a medicament in the treatment of colon carcinoma.

36. Use of a polypeptide as claimed in any one of claims 1 to 7 for the manufacture of a medicament in the treatment of carcinoma.

37. Use of a polypeptide as claimed in any one of claims 1 to 7 for the manufacture of a medicament in the treatment of colon carcinoma.

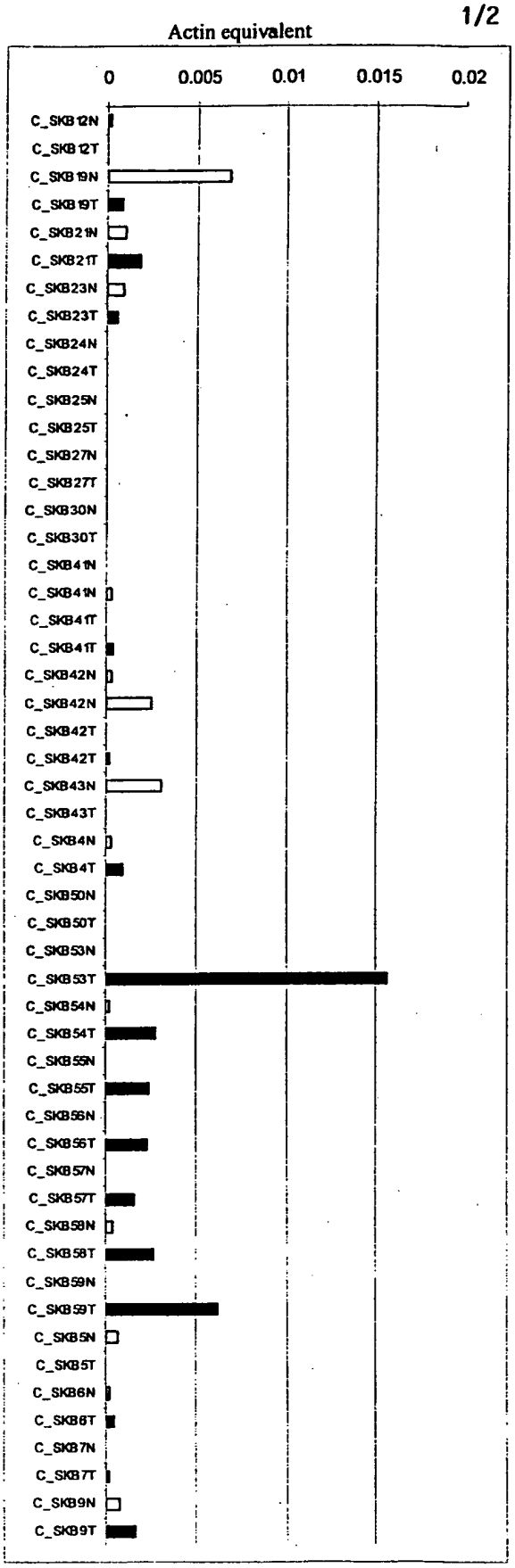

Figure 1: Real-time PCR data on matched normal and tumoral colon samples

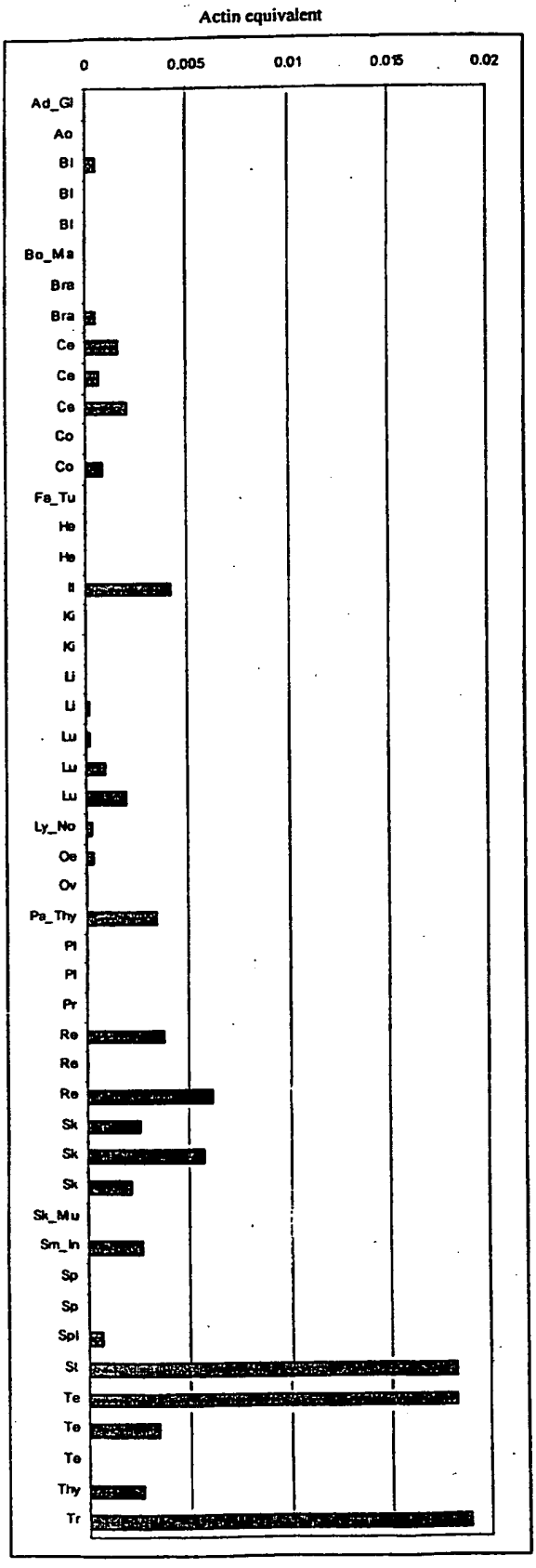

Figure 2: Real-time PCR data of CASB619 in normal tissues

Legend: Ad_Gl: adrenal gland; Ao: aorta; Bl: bladder, Bo_Ma: bone marrow; Bra: brain; Ce: cervix; Co: colon; Fa_Tu: fallopian tube; He: heart; Il: ileon; Ki: kidney; Li: liver; Lu: lung; Ly_No: lymph node; Oe: oesophagus; Ov: ovary; a_Thy: parathyroid gland; Pl: placenta; Pr: prostate; Re: rectum; Sk: skin; Sk_Mu: skeletal muscle; Sp: pleen; St: stomach; Te: testis; Thy: thyroid gland; Tr: trachea

SEQUENCE LISTING

<110> SmithKline Beecham Biologicals S.A.

<120> Novel Compounds

<130> BC45226

<160> 68

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 3280
<212> DNA
<213> homo sapiens

<400> 1

atggctgagc ctgggcacag ccaccatctc tccgccagag tcaggggaag aactgagagg 60
cgcataccc ggctgtggcg gctgctgctc tgggctggga ccgccttcca ggtgacccag 120
ggaacgggac cggagcttca tgcctgcaaa gagtctgagt accactatga gtacacggcg 180
tgtgacagca cgggttccag gtggagggtc gccgtgccgc ataccccggg cctgtgcacc 240
agcctgcctg accccgtcaa gggcaccgag tgctccttct cctgcaacgc cggggagttt 300
ctggatatga aggaccagtc atgtaagcca tgcgctgagg gccgctactc cctcggcaca 360
ggcattcggt ttgatgagtg ggatgagctg ccccatggct ttgccagcct ctcagccaac 420
atggagctgg atgacagtgc tgctgagtcc accgggaact gtacttcgtc caagtgggtt 480
ccccggggcg actacatcgc ctccaacacg gacgaatgca cagccacact gatgtacgcc 540
gtcaacctga agcaatctgg caccgttaac ttcgaatact actatccaga ctccagcatc 600
atctttgagt ttttcgttca gaatgaccag tgccagccca atgcagatga ctccaggtgg 660
atgaagacca cagagaaagg atgggaattc cacagtgtgg agctaaatcg aggcaataat 720
gtcctctatt ggagaaccac agccttctca gtatggacca aagtacccaa gcctgtgctg 780
gtgagaaaca ttgccataac aggggtggcc tacacttcag aatgcttccc ctgcaaacct 840
ggcacgtatg cagacaagca gggctcctct ttctgcaaac tttgcccagc caactcttat 900
tcaaataaag gagaaacttc ttgccaccag tgtgaccctg acaaatactc agagaaagga 960
tcttcttcct gtaacgtgcg cccagcttgc acagacaaag attatttcta cacacacacg 1020
gcctgcgatg ccaacggaga gacacaactc atgtacaaat gggccaagcc gaaaatctgt 1080
agcgaggacc ttgagggggc agtgaagctg cctgcctctg gtgtgaagac ccactgccca 1140
ccctgcaacc caggcttctt caaaaccaac aacagcacct gccagccctg cccatatggt 1200
tcctactcca atggctcaga ctgtacccgc tgccctgcag ggactgaacc tgctgtggga 1260
tttgaataca aatggtggaa cacgctgccc acaaacatgg aaacgaccgt tctcagtggg 1320
atcaacttcg agtacaaggg catgacaggc tgggaggtgg ctggtgatca catttacaca 1380
gctgctggag cctcagacaa tgacttcatg attctcactc tggttgtgcc aggatttaga 1440
cctccgcagt cggtgatggc agacacagag aataaagagg tggccagaat cacatttgtc 1500
tttgagaccc tctgttctgt gaactgtgag ctctacttca tggtgggtgt gaattctagg 1560
accaacactc ctgtggagac gtggaaaggt tccaaaggca aacagtccta tacctacatc 1620
attgaggaga acactaccac gagcttcacc tgggccttcc agaggaccac ttttcatgag 1680
gcaagcagga agtacaccaa tgacgttgcc aagatctact ccatcaatgt caccaatgtt 1740
atgaatggcg tggcctccta ctgccgtccc tgtgccctag aagcctctga tgtgggctcc 1800
tcctgcacct cttgtcctgc tggttactat attgaccgag attcaggaac ctgccactcc 1860
tgccccccta acacaattct gaaagcccac cagccttatg gtgtccaggc ctgtgtgccc 1920
tgtggtccag ggaccaagaa caacaagatc cactctctgt gctacaatga ttgcaccttc 1980
tcacgcaaca ctccaaccag gactttcaac tacaacttct ccgctttggc aaacaccgtc 2040
actcttgctg gagggccaag cttcacttcc aaagggttga aatacttcca tcactttacc 2100
ctcagtctct gtggaaacca gggtaggaaa atgtctgtgt gcaccgacaa tgtcactgac 2160
ctccggattc ctgagggtga gtcagggttc tccaaatcta tcacagccta cgtctgccag 2220
gcagtcatca tccccccaga ggtgacaggc tacaaggccg gggtttcctc acagcctgtc 2280
agccttgctg atcgacttat tggggtgaca acagatatga ctctggatgg aatcacctcc 2340
ccagctgaac ttttccacct ggagtccttg ggaataccgg acgtgatctt cttttatagg 2400
tccaatgatg tgacccagtc ctgcagttct gggagatcaa ccaccatccg cgtcaggtgc 2460
agtccacaga aaactgtccc tggaagtttg ctgctgccag gaacgtgctc agatgggacc 2520
tgtgatggct gcaacttcca cttcctgtgg gagagcgcgg ctgcttgccc gctctgctca 2580

```
gtggctgact accatgctat cgtcagcagc tgtgtggctg ggatccagaa gactacttac    2640
gtgtggcgag aacccaagct atgctctggt ggcatttctc tgcctgagca gagagtcacc    2700
atctgcaaaa ccatagattt ctggctgaaa gtgggcatct ctgcaggcac ctgtactgcc    2760
atcctgctca ccgtcttgac ctgctacttt tggaaaaaga atcaaaaact agagtacaag    2820
tactccaagc tggtgatgaa tgctactctc aaggactgtg acctgccagc agctgacagc    2880
tgcgccatca tggaaggcga ggatgtagag gacgacctca tctttaccag caagaagtca    2940
ctctttggga agatcaaatc atttacctcc aagaggactc ctgatggatt tgactcagtg    3000
ccgctgaaga catcctcagg aggcccagac atggacctgt gagaggcact gcctgcctca    3060
cctgcctcct caccttgcat agcacctttg caagcctgcg gcgatttggg tgccagcatc    3120
ctgcaacacc cactgctgga aatctcttca ttgtggcctt atcagatgtt tgaatttcag    3180
atctttttt atagagtacc caaaccctcc tttctgcttg cctcaaacct gccaaatata    3240
cccacacttt gtttgtaaaa aaaaaaaaaa aaaaaaaaaa                          3280
```

<210> 2
<211> 1013
<212> PRT
<213> homo sapiens

<400> 2

```
Met Ala Glu Pro Gly His Ser His His Leu Ser Ala Arg Val Arg Gly
 1               5                  10                  15
Arg Thr Glu Arg Arg Ile Pro Arg Leu Trp Arg Leu Leu Leu Trp Ala
            20                  25                  30
Gly Thr Ala Phe Gln Val Thr Gln Gly Thr Gly Pro Glu Leu His Ala
        35                  40                  45
Cys Lys Glu Ser Glu Tyr His Tyr Glu Tyr Thr Ala Cys Asp Ser Thr
    50                  55                  60
Gly Ser Arg Trp Arg Val Ala Val Pro His Thr Pro Gly Leu Cys Thr
65                  70                  75                  80
Ser Leu Pro Asp Pro Val Lys Gly Thr Glu Cys Ser Phe Ser Cys Asn
                85                  90                  95
Ala Gly Glu Phe Leu Asp Met Lys Asp Gln Ser Cys Lys Pro Cys Ala
            100                 105                 110
Glu Gly Arg Tyr Ser Leu Gly Thr Gly Ile Arg Phe Asp Glu Trp Asp
        115                 120                 125
Glu Leu Pro His Gly Phe Ala Ser Leu Ser Ala Asn Met Glu Leu Asp
    130                 135                 140
Asp Ser Ala Ala Glu Ser Thr Gly Asn Cys Thr Ser Ser Lys Trp Val
145                 150                 155                 160
Pro Arg Gly Asp Tyr Ile Ala Ser Asn Thr Asp Glu Cys Thr Ala Thr
                165                 170                 175
Leu Met Tyr Ala Val Asn Leu Lys Gln Ser Gly Thr Val Asn Phe Glu
            180                 185                 190
Tyr Tyr Tyr Pro Asp Ser Ser Ile Ile Phe Glu Phe Phe Val Gln Asn
        195                 200                 205
Asp Gln Cys Gln Pro Asn Ala Asp Asp Ser Arg Trp Met Lys Thr Thr
    210                 215                 220
Glu Lys Gly Trp Glu Phe His Ser Val Glu Leu Asn Arg Gly Asn Asn
225                 230                 235                 240
Val Leu Tyr Trp Arg Thr Thr Ala Phe Ser Val Trp Thr Lys Val Pro
                245                 250                 255
Lys Pro Val Leu Val Arg Asn Ile Ala Ile Thr Gly Val Ala Tyr Thr
            260                 265                 270
Ser Glu Cys Phe Pro Cys Lys Pro Gly Thr Tyr Ala Asp Lys Gln Gly
        275                 280                 285
Ser Ser Phe Cys Lys Leu Cys Pro Ala Asn Ser Tyr Ser Asn Lys Gly
    290                 295                 300
Glu Thr Ser Cys His Gln Cys Asp Pro Asp Lys Tyr Ser Glu Lys Gly
305                 310                 315                 320
Ser Ser Ser Cys Asn Val Arg Pro Ala Cys Thr Asp Lys Asp Tyr Phe
                325                 330                 335
Tyr Thr His Thr Ala Cys Asp Ala Asn Gly Glu Thr Gln Leu Met Tyr
```

```
                340                 345                 350
Lys Trp Ala Lys Pro Lys Ile Cys Ser Glu Asp Leu Glu Gly Ala Val
            355                 360                 365
Lys Leu Pro Ala Ser Gly Val Lys Thr His Cys Pro Pro Cys Asn Pro
        370                 375                 380
Gly Phe Phe Lys Thr Asn Asn Ser Thr Cys Gln Pro Cys Pro Tyr Gly
385                 390                 395                 400
Ser Tyr Ser Asn Gly Ser Asp Cys Thr Arg Cys Pro Ala Gly Thr Glu
                405                 410                 415
Pro Ala Val Gly Phe Glu Tyr Lys Trp Trp Asn Thr Leu Pro Thr Asn
            420                 425                 430
Met Glu Thr Thr Val Leu Ser Gly Ile Asn Phe Glu Tyr Lys Gly Met
        435                 440                 445
Thr Gly Trp Glu Val Ala Gly Asp His Ile Tyr Thr Ala Ala Gly Ala
    450                 455                 460
Ser Asp Asn Asp Phe Met Ile Leu Thr Leu Val Val Pro Gly Phe Arg
465                 470                 475                 480
Pro Pro Gln Ser Val Met Ala Asp Thr Glu Asn Lys Glu Val Ala Arg
                485                 490                 495
Ile Thr Phe Val Phe Glu Thr Leu Cys Ser Val Asn Cys Glu Leu Tyr
            500                 505                 510
Phe Met Val Gly Val Asn Ser Arg Thr Asn Thr Pro Val Glu Thr Trp
        515                 520                 525
Lys Gly Ser Lys Gly Lys Gln Ser Tyr Thr Tyr Ile Ile Glu Glu Asn
    530                 535                 540
Thr Thr Thr Ser Phe Thr Trp Ala Phe Gln Arg Thr Thr Phe His Glu
545                 550                 555                 560
Ala Ser Arg Lys Tyr Thr Asn Asp Val Ala Lys Ile Tyr Ser Ile Asn
                565                 570                 575
Val Thr Asn Val Met Asn Gly Val Ala Ser Tyr Cys Arg Pro Cys Ala
            580                 585                 590
Leu Glu Ala Ser Asp Val Gly Ser Ser Cys Thr Ser Cys Pro Ala Gly
        595                 600                 605
Tyr Tyr Ile Asp Arg Asp Ser Gly Thr Cys His Ser Cys Pro Pro Asn
    610                 615                 620
Thr Ile Leu Lys Ala His Gln Pro Tyr Gly Val Gln Ala Cys Val Pro
625                 630                 635                 640
Cys Gly Pro Gly Thr Lys Asn Asn Lys Ile His Ser Leu Cys Tyr Asn
                645                 650                 655
Asp Cys Thr Phe Ser Arg Asn Thr Pro Thr Arg Thr Phe Asn Tyr Asn
            660                 665                 670
Phe Ser Ala Leu Ala Asn Thr Val Thr Leu Ala Gly Gly Pro Ser Phe
        675                 680                 685
Thr Ser Lys Gly Leu Lys Tyr Phe His His Phe Thr Leu Ser Leu Cys
    690                 695                 700
Gly Asn Gln Gly Arg Lys Met Ser Val Cys Thr Asp Asn Val Thr Asp
705                 710                 715                 720
Leu Arg Ile Pro Glu Gly Glu Ser Gly Phe Ser Lys Ser Ile Thr Ala
                725                 730                 735
Tyr Val Cys Gln Ala Val Ile Ile Pro Pro Glu Val Thr Gly Tyr Lys
            740                 745                 750
Ala Gly Val Ser Ser Gln Pro Val Ser Leu Ala Asp Arg Leu Ile Gly
        755                 760                 765
Val Thr Thr Asp Met Thr Leu Asp Gly Ile Thr Ser Pro Ala Glu Leu
    770                 775                 780
Phe His Leu Glu Ser Leu Gly Ile Pro Asp Val Ile Phe Phe Tyr Arg
785                 790                 795                 800
Ser Asn Asp Val Thr Gln Ser Cys Ser Ser Gly Arg Ser Thr Thr Ile
                805                 810                 815
Arg Val Arg Cys Ser Pro Gln Lys Thr Val Pro Gly Ser Leu Leu Leu
            820                 825                 830
Pro Gly Thr Cys Ser Asp Gly Thr Cys Asp Gly Cys Asn Phe His Phe
```

```
         835                 840                 845
Leu Trp Glu Ser Ala Ala Ala Cys Pro Leu Cys Ser Val Ala Asp Tyr
     850                 855                 860
His Ala Ile Val Ser Ser Cys Val Ala Gly Ile Gln Lys Thr Thr Tyr
865                 870                 875                 880
Val Trp Arg Glu Pro Lys Leu Cys Ser Gly Gly Ile Ser Leu Pro Glu
                885                 890                 895
Gln Arg Val Thr Ile Cys Lys Thr Ile Asp Phe Trp Leu Lys Val Gly
            900                 905                 910
Ile Ser Ala Gly Thr Cys Thr Ala Ile Leu Leu Thr Val Leu Thr Cys
        915                 920                 925
Tyr Phe Trp Lys Lys Asn Gln Lys Leu Glu Tyr Lys Tyr Ser Lys Leu
    930                 935                 940
Val Met Asn Ala Thr Leu Lys Asp Cys Asp Leu Pro Ala Ala Asp Ser
945                 950                 955                 960
Cys Ala Ile Met Glu Gly Glu Asp Val Glu Asp Asp Leu Ile Phe Thr
                965                 970                 975
Ser Lys Lys Ser Leu Phe Gly Lys Ile Lys Ser Phe Thr Ser Lys Arg
            980                 985                 990
Thr Pro Asp Gly Phe Asp Ser Val Pro Leu Lys Thr Ser Ser Gly Gly
        995                 1000                1005
Pro Asp Met Asp Leu
    1010
```

<210> 3
<211> 677
<212> DNA
<213> homo sapiens

<400> 3

```
ttttttaatt tacaaacaaa gtgtgggtat atttggcagg tttgaggcaa gcagaaagga     60
gggtttgggt actctataaa aaaagatctg aaattcaaac atctgataag gccacaatga    120
agagatttcc agcagtgggt gttgcaggat gctggcaccc aaatcgccgc acgttgcaaa    180
ggtgctatgc aaggtgagga ggcaggtgag gcaggcagtg cctctcacag gtccatgtct    240
gggcctcctg aggatgtctt cagcggcact gagtcaaatc catcaggagt cctcttggag    300
gtaaatgatt tgatcttccc aaagagtgac ttcttgctgg taaagatgag gtcgtcctct    360
acatcctcgc cttccatgat ggcgcagtgt cagctgctgg caggtcacag tccttgagag    420
tagcattcat caccagcttg gagtacttgt actctagttt ttgattcttt ttccaaaagt    480
agcaggtcaa gacggtgagc aggatggcag tacaggtgcc tgcagagatg cccactttca    540
gccagaaatc tatggttttg cagatggtga ctctctgctc aggcagagaa atgccaccag    600
agcatagctt gggttctcgc cacacgtaag tagtcttctg gatcccagcc acacagctgc    660
tgacgatagc atggtag                                                   677
```

<210> 4
<211> 105
<212> PRT
<213> homo sapiens

<400> 4

```
Tyr His Ala Ile Val Ser Ser Cys Val Ala Gly Ile Gln Lys Thr Thr
 1               5                  10                  15
Tyr Val Trp Arg Glu Pro Lys Leu Cys Ser Gly Gly Ile Ser Leu Pro
            20                  25                  30
Glu Gln Arg Val Thr Ile Cys Lys Thr Ile Asp Phe Trp Leu Lys Val
        35                  40                  45
Gly Ile Ser Ala Gly Thr Cys Thr Ala Ile Leu Leu Thr Val Leu Thr
    50                  55                  60
Cys Tyr Phe Trp Lys Lys Asn Gln Lys Leu Glu Tyr Lys Tyr Ser Lys
65                  70                  75                  80
Leu Val Met Asn Ala Thr Leu Lys Asp Cys Asp Leu Pro Ala Ala Asp
                85                  90                  95
Thr Ala Pro Ser Trp Lys Ala Arg Met
```

100 105

<210> 5
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 5
Arg Leu Trp Arg Leu Leu Leu Trp Ala
1 5

<210> 6
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 6
Ser Leu Ala Asp Arg Leu Ile Gly Val
1 5

<210> 7
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 7
Ser Leu Pro Glu Gln Arg Val Thr Ile
1 5

<210> 8
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 8
Lys Leu Cys Ser Gly Gly Ile Ser Leu
1 5

<210> 9
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 9
Ala Ala Ala Cys Pro Leu Cys Ser Val
1 5

<210> 10
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 10
Ser Ala Leu Ala Asn Thr Val Thr Leu
1 5

<210> 11
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 11

Phe Val Phe Glu Thr Leu Cys Ser Val
1 5

<210> 12
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 12
Glu Leu Pro His Gly Phe Ala Ser Leu
1 5

<210> 13
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 13
Leu Ile Phe Thr Ser Lys Lys Ser Leu
1 5

<210> 14
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 14
Lys Leu Glu Tyr Lys Tyr Ser Lys Leu
1 5

<210> 15
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 15
Lys Thr Ile Asp Phe Trp Leu Lys Val
1 5

<210> 16
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 16
Ser Val Ala Asp Tyr His Ala Ile Val
1 5

<210> 17
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 17
Leu Leu Leu Pro Gly Thr Cys Ser Asp
1 5

<210> 18
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 18
Ala Leu Ala Asn Thr Val Thr Leu Ala
1 5

<210> 19
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 19
Thr Leu Cys Ser Val Asn Cys Glu Leu
1 5

<210> 20
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 20
Asn Thr Asp Glu Cys Thr Ala Thr Leu
1 5

<210> 21
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 21
Ser Leu Pro Asp Pro Val Lys Gly Thr
1 5

<210> 22
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 22
Ser Leu Phe Gly Lys Ile Lys Ser Phe
1 5

<210> 23
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 23
Cys Thr Ala Ile Leu Leu Thr Val Leu
1 5

<210> 24
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 24
Ile Val Ser Ser Cys Val Ala Gly Ile
1 5

<210> 25
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 25
Lys Met Ser Val Cys Thr Asp Asn Val
1 5

<210> 26
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 26
Val Leu Val Arg Asn Ile Ala Ile Thr
1 5

<210> 27
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 27
Glu Leu Asn Arg Gly Asn Asn Val Leu
1 5

<210> 28
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 28
Ala Thr Leu Met Tyr Ala Val Asn Leu
1 5

<210> 29
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 29
Val Ala Val Pro His Thr Pro Gly Leu
1 5

<210> 30
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 30
Arg Leu Trp Arg Leu Leu Leu Trp Ala
1 5

<210> 31
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 31
Ser Ala Gly Thr Cys Thr Ala Ile Leu
1 5

<210> 32
<211> 9
<212> PRT

<213> Artificial Sequence

<400> 32
Gly Ile Ser Leu Pro Glu Gln Arg Val
1 5

<210> 33
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 33
Lys Thr Val Pro Gly Ser Leu Leu Leu
1 5

<210> 34
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 34
Arg Leu Ile Gly Val Thr Thr Asp Met
1 5

<210> 35
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 35
Thr Leu Ala Gly Gly Pro Ser Phe Thr
1 5

<210> 36
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 36
Tyr Ile Ile Glu Glu Asn Thr Thr Thr
1 5

<210> 37
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 37
Ile Ala Ile Thr Gly Val Ala Tyr Thr
1 5

<210> 38
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 38
Val Thr Gln Gly Thr Gly Pro Glu Leu
1 5

<210> 39
<211> 10

<212> PRT
<213> Artificial Sequence

<400> 39
Ser Leu Phe Gly Lys Ile Lys Ser Phe Thr
1 5 10

<210> 40
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 40
Ala Ile Val Ser Ser Cys Val Ala Gly Ile
1 5 10

<210> 41
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 41
Ser Ala Ala Ala Cys Pro Leu Cys Ser Val
1 5 10

<210> 42
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 42
His Leu Glu Ser Leu Gly Ile Pro Asp Val
1 5 10

<210> 43
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 43
Lys Ile Tyr Ser Ile Asn Val Thr Asn Val
1 5 10

<210> 44
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 44
Ser Leu Ala Asp Arg Leu Ile Gly Val Thr
1 5 10

<210> 45
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 45
Ile Leu Lys Ala His Gln Pro Tyr Gly Val
1 5 10

<210> 46

<211> 10
<212> PRT
<213> Artificial Sequence

<400> 46
Val Met Ala Asp Thr Glu Asn Lys Glu Val
1 5 10

<210> 47
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 47
Leu Leu Trp Ala Gly Thr Ala Phe Gln Val
1 5 10

<210> 48
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 48
Gly Thr Cys Thr Ala Ile Leu Leu Thr Val
1 5 10

<210> 49
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 49
Ile Thr Ser Pro Ala Glu Leu Phe His Leu
1 5 10

<210> 50
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 50
Leu Ile Gly Val Thr Thr Asp Met Thr Leu
1 5 10

<210> 51
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 51
Thr Leu Pro Thr Asn Met Glu Thr Thr Val
1 5 10

<210> 52
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 52
Leu Met Tyr Lys Trp Ala Lys Pro Lys Ile
1 5 10

<210> 53
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 53
Asp Leu Ile Phe Thr Ser Lys Lys Ser Leu
1 5 10

<210> 54
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 54
Gly Leu Lys Tyr Phe His His Phe Thr Leu
1 5 10

<210> 55
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 55
Gly Thr Lys Asn Asn Lys Ile His Ser Leu
1 5 10

<210> 56
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 56
Ser Asp Asn Asp Phe Met Ile Leu Thr Leu
1 5 10

<210> 57
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 57
Leu Val Arg Asn Ile Ala Ile Thr Gly Val
1 5 10

<210> 58
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 58
Gly Leu Cys Thr Ser Leu Pro Asp Pro Val
1 5 10

<210> 59
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 59
Thr Leu Lys Asp Cys Asp Leu Pro Ala Ala
1 5 10

<210> 60
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 60
Val Ile Phe Phe Tyr Arg Ser Asn Asp Val
1 5 10

<210> 61
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 61
Ser Ile Thr Ala Tyr Val Cys Gln Ala Val
1 5 10

<210> 62
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 62
Ser Val Cys Thr Asp Asn Val Thr Asp Leu
1 5 10

<210> 63
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 63
Ser Leu Cys Gly Asn Gln Gly Arg Lys Met
1 5 10

<210> 64
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 64
Asn Met Glu Thr Thr Val Leu Ser Gly Ile
1 5 10

<210> 65
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 65
Asn Ile Ala Ile Thr Gly Val Ala Tyr Thr
1 5 10

<210> 66
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 66
Gly Ile Arg Phe Asp Glu Trp Asp Glu Leu

1 5 10

<210> 67
<211> 10
<212> PRT
<213> Artificial Sequence

<400> 67
Phe His Phe Leu Trp Glu Ser Ala Ala Ala
1 5 10

<210> 68
<211> 9
<212> PRT
<213> Artificial Sequence

<400> 68
Phe Leu Trp Glu Ser Ala Ala Ala Cys
1 5

SEQUENCE INFORMATION

SEQ ID NO:1

atggctgagcctgggcacagccaccatctctccgccagagtcaggggaagaactgagaggcgcataccccggctgtggcg
gctgctgctctgggctgggaccgccttccaggtgacccagggaacgggaccggagcttcatgcctgcaaagagtctgagt
accactatgagtacacggcgtgtgacagcacgggttccaggtggagggtcgccgtgccgcataccccgggcctgtgcacc
agcctgcctgaccccgtcaagggcaccgagtgctccttctcctgcaacgccggggagtttctggatatgaaggaccagtc
atgtaagccatgcgctgagggccgctactccctcggcacaggcattcggtttgatgagtgggatgagctgccccatggct
ttgccagcctctcagccaacatggagctggatgacagtgctgctgagtccaccgggaactgtacttcgtccaagtgggtt
ccccggggcgactacatcgcctccaacacggacgaatgcacagccacactgatgtacgccgtcaacctgaagcaatctgg
caccgttaacttcgaatactactatccagactccagcatcatctttgagtttttcgttcagaatgaccagtgccagccca
atgcagatgactccaggtggatgaagaccacagagaaaggatgggaattccacagtgtggagctaaatcgaggcaataat
gtcctctattggagaaccacagccttctcagtatggaccaaagtacccaagcctgtgctggtgagaaacattgccataac
aggggtggcctacacttcagaatgcttcccctgcaaacctggcacgtatgcagacaagcagggctcctctttctgcaaac
tttgcccagccaactcttattcaaataaaggagaaacttcttgccaccagtgtgaccctgacaaatactcagagaaagga
tcttcttcctgtaacgtgcgcccagcttgcacagacaaagattatttctacacacacacggcctgcgatgccaacggaga
gacacaactcatgtacaaatgggccaagccgaaaatctgtagcgaggaccttgagggggcagtgaagctgcctgcctctg
gtgtgaagacccactgcccaccctgcaacccaggcttcttcaaaaccaacaacagcacctgccagccctgcccatatggt
tcctactccaatggctcagactgtacccgctgccctgcagggactgaacctgctgtgggatttgaatacaaatggtggaa
cacgctgcccacaaacatggaaacgaccgttctcagtgggatcaacttcgagtacaagggcatgacaggctgggaggtgg
ctggtgatcacatttacacagctgctggagcctcagacaatgacttcatgattctcactctggttgtgccaggatttaga
cctccgcagtcggtgatggcagacacagagaataaagaggtggccagaatcacatttgtctttgagaccctctgttctgt
gaactgtgagctctacttcatggtgggtgtgaattctaggaccaacactcctgtggagacgtggaaaggttccaaaggca
aacagtcctatacctacatcattgaggagaacactaccacgagcttcacctgggccttccagaggaccacttttcatgag
gcaagcaggaagtacaccaatgacgttgccaagatctactccatcaatgtcaccaatgttatgaatggcgtggcctccta
ctgccgtccctgtgccctagaagcctctgatgtgggctcctcctgcacctcttgtcctgctggttactatattgaccgag
attcaggaacctgccactcctgccccctaacacaattctgaaagcccaccagccttatggtgtccaggcctgtgtgccc
tgtggtccagggaccaagaacaacaagatccactctctgtgctacaatgattgcaccttctcacgcaacactccaaccag
gactttcaactacaacttctccgctttggcaaacaccgtcactcttgctggagggccaagcttcacttccaaagggttga
aatacttccatcactttaccctcagtctctgtggaaaccagggtaggaaaatgtctgtgtgcaccgacaatgtcactgac
ctccggattcctgagggtgagtcagggttctccaaatctatcacagcctacgtctgccaggcagtcatcatccccccaga
ggtgacaggctacaaggccggggtttcctcacagcctgtcagccttgctgatcgacttattggggtgacaacagatatga
ctctggatggaatcacctccccagctgaacttttccacctggagtccttgggaataccggacgtgatcttcttttatagg
tccaatgatgtgacccagtcctgcagttctgggagatcaaccaccatccgcgtcaggtgcagtccacagaaaactgtccc
tggaagtttgctgctgccaggaacgtgctcagatgggacctgtgatggctgcaacttccacttcctgtgggagagcgcgg
ctgcttgcccgctctgctcagtggctgactaccatgctatcgtcagcagctgtgtggctgggatccagaagactacttac
gtgtggcgagaacccaagctatgctctggtggcatttctctgcctgagcagagagtcaccatctgcaaaaccatagattt
ctggctgaaagtgggcatctctgcaggcacctgtactgccatcctgctcaccgtcttgacctgctacttttggaaaaaga
atcaaaaactagagtacaagtactccaagctggtgatgaatgctactctcaaggactgtgacctgccagcagctgacagc
tgcgccatcatggaaggcgaggatgtagaggacgacctcatctttaccagcaagaagtcactctttgggaagatcaaatc
atttacctccaagaggactcctgatggatttgactcagtgccgctgaagacatcctcaggaggcccagacatggacctgt

gaGAGGCACTGCCTGCCTCACCTGCCTCCTCACCTTGCATAGCACCTTTGCAAGCCTGCGGCGATTTGGGTGCCAGCATC
CTGCAACACCCACTGCTGGAAATCTCTTCATTGTGGCCTTATCAGATGTTTGAATTTCAGATCTTTTTTTATAGAGTACC
CAAACCCTCCTTTCTGCTTGCCTCAAACCTGCCAAATATACCCACACTTTGTTTGTAAAAaaaAAAAAAAAAAAAAAAAA

SEQ ID NO:2

MAEPGHSHHLSARVRGRTERRIPRLWRLLLWAGTAFQVTQGTGPELHACKESEYHYEYTACDSTGSRWRVAVPHTPGLCT
SLPDPVKGTECSFSCNAGEFLDMKDQSCKPCAEGRYSLGTGIRFDEWDELPHGFASLSANMELDDSAAESTGNCTSSKWV
PRGDYIASNTDECTATLMYAVNLKQSGTVNFEYYYPDSSIIFEFFVQNDQCQPNADDSRWMKTTEKGWEFHSVELNRGNN
VLYWRTTAFSVWTKVPKPVLVRNIAITGVAYTSECFPCKPGTYADKQGSSFCKLCPANSYSNKGETSCHQCDPDKYSEKG
SSSCNVRPACTDKDYFYTHTACDANGETQLMYKWAKPKICSEDLEGAVKLPASGVKTHCPPCNPGFFKTNNSTCQPCPYG
SYSNGSDCTRCPAGTEPAVGFEYKWWNTLPTNMETTVLSGINFEYKGMTGWEVAGDHIYTAAGASDNDFMILTLVVPGFR
PPQSVMADTENKEVARITFVFETLCSVNCELYFMVGVNSRTNTPVETWKGSKGKQSYTYIIEENTTTSFTWAFQRTTFHE
ASRKYTNDVAKIYSINVTNVMNGVASYCRPCALEASDVGSSCTSCPAGYYIDRDSGTCHSCPPNTILKAHQPYGVQACVP
CGPGTKNNKIHSLCYNDCTFSRNTPTRTFNYNFSALANTVTLAGGPSFTSKGLKYFHHFTLSLCGNQGRKMSVCTDNVTD
LRIPEGESGFSKSITAYVCQAVIIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGITSPAELFHLESLGIPDVIFFYR
SNDVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPGTCSDGTCDGCNFHFLWESAAACPLCSVADYHAIVSSCVAGIQKTTY
VWREPKLCSGGISLPEQRVTICKTIDFWLKVGISAGTCTAILLTVLTCYFWKKNQKLEYKYSKLVMNATLKDCDLPAADS
CAIMEGEDVEDDLIFTSKKSLFGKIKSFTSKRTPDGFDSVPLKTSSGGPDMDL

SEQ ID NO:3

TTTTTTAATTTACAAACAAAGTGTGGGTATATTTGGCAGGTTTGAGGCAAGCAGAAAGGAGGGTTTGGGTAC
TCTATAAAAAAAGATCTGAAATTCAAACATCTGATAAGGCCACAATGAAGAGATTTCCAGCAGTGGGTGTTG
CAGGATGCTGGCACCCAAATCGCCGCACGTTGCAAAGGTGCTATGCAAGGTGAGGAGGCAGGTGAGGCAGGC
AGTGCCTCTCACAGGTCCATGTCTGGGCCTCCTGAGGATGTCTTCAGCGGCACTGAGTCAAATCCATCAGGA
GTCCTCTTGGAGGTAAATGATTTGATCTTCCCAAAGAGTGACTTCTTGCTGGTAAAGATGAGGTCGTCCTCT
ACATCCTCGCCTTCCATGATGGCGCAGTGTCAGCTGCTGGCAGGTCACAGTCCTTGAGAGTAGCATTCATCA
CCAGCTTGGAGTACTTGTACTCTAGTTTTTGATTCTTTTTCCAAAAGTAGCAGGTCAAGACGGTGAGCAGGA
TGGCAGTACAGGTGCCTGCAGAGATGCCCACTTTCAGCCAGAAATCTATGGTTTTGCAGATGGTGACTCTCT
GCTCAGGCAGAGAAATGCCACCAGAGCATAGCTTGGGTTCTCGCCACACGTAAGTAGTCTTCTGGATCCCAG
CCACACAGCTGCTGACGATAGCATGGTAG

SEQ ID NO:4

YHAIVSSCVAGIQKTTYVWREPKLCSGGISLPEQRVTICKTIDFWLKVGISAGTCTAILLTVLTCYF
WKKNQKLEYKYSKLVMNATLKDCDLPAADTAPSWKARM